Journal of Multimedia

Contents

A Novel Binocular Vision System for Surveillance Application

Zhigao Cui and Aihua Li
502 Faculty, Xi'an Institute of High-tech, Xi'an, China
Email: cuizg10@126.com, LCC@126.com

Abstract—**Visual surveillance system has been well studied in past decades. Because of the conflict between surveillance range and resolution of single-camera system, visual surveillance using multiple cameras has attracted increasing interest in recent years. In multi-camera systems, dual-camera system which contains active camera such as PTZ (pan-tilt-zoom) camera is the simplest and typical one. The superiority of this system lies in that it can obtain multi-resolution information and depth information. But as far as we know, few of the dual systems make full use of the two superiorities. In this paper, we propose a new binocular vision system which is composed of two PTZ cameras. Different from other systems, stereo rectification is used to establish correspondence between two image coordinates. Then, both global and detailed image information can be obtained. Furthermore, we use the depth information to help solve the occlusion problem in tracking. The experimental results on indoor surveillance environments have demonstrated the effectiveness and robustness of our system.**

Index Terms—**Binocular Vision System, PTZ (Pan-Tilt-Zoom) Camera, Stereo Rectification, Occlusion**

I. INTRODUCTION

With the rapid demands of security and safety, vision-based surveillance technologies have become an active subject in the field of computer vision [1-3]. Some large research projects have been conducted in many countries, e.g., the VSAM project supported by DARPA, the ESPRIT PASSWORDS project in Europe and Japan's Cooperative Distributed Vision project. These systems have been well applied in commercial, law enforcement and military fields. The aim of vision-based surveillance system is to replace traditional passive surveillance system and accomplish the entire surveillance task as automatically as possible. There are several important factors in visual surveillance: global information, specific attention with high resolution (local information), stereo information and so on.

According to the number of cameras, the visual surveillance system can be classified into monocular and multi-camera system. Systems using single camera (monocular) have been widely applied in the past decades [4-6]. However, they are limited by their fixed view angles, fixed resolutions and limited depth information. Due to rapid progress in computer vision, camera control, network cooperation [7] technology, hardware development, and so on, multi-camera systems'

superiority become more and more obvious. Binocular vision system is a simple and typical one. In the past decades, lots of related works have been published. We only review some of them which we consider typical. Ref. [8] proposes a vision system using two omnidirectional cameras. This type of vision system is mainly used for robot vision, which can enlarge view field. Ref. [9] proposes a vision system with one omnidirectional camera and one PTZ camera. This kind of vision system is mostly applied in indoor environment. The omnidirectional camera is used for monitoring the whole surveillance scene, and PTZ camera is used for acquiring high resolution image of interested target. The two cameras work in a fixed master-salve mode. The main disadvantage of the above vision systems is the usage of the omnidirectional camera, whose image resolution is low and uneven. This drawback has great limitation in real application.

In addition to binocular vision system mentioned above, dual static, static plus active or dual active system are the mostly well applied vision systems in real application. Traditional dual static system could get depth information by stereo vision. Generally speaking, the larger the baseline, the more difficult the correspondence between two image points can be build. This kind of system can be integrated as a vision unit to improve its stability. All these systems work on a symmetrical mode. Ref. [10] proposes an object tracking method that combines color and depth information using dual static cameras. The works in [11] propose that use of plan-view maps to represent stereo information more efficiently. The main drawback of this kind of systems is that it could not obtain multi-resolution and multi-visual-angle information. Static plus active or dual active system always works on master-slave mode, which means that the two cameras are asymmetrical. In [12], Zhou et al. present a master-slave system for acquiring biometric imagery of interesting subject in an outdoor environment. The relationship between static image coordinates and active camera's PTZ parameters is contained in a sample-interpolated mapping. Ref. [13] is similar to Ref. [12] in that it is also for outdoor scene, but it assumes that the ground in the area can be viewed as planar, and a linear transform associate with a mapping between object height in the master image and height in the slave image is applied to align two images. Bodor et al. present in Ref. [14] a system for both indoor and outdoor scene. Their

system disregards the base line between two cameras, and uses planar mapping from pixel in static image to pan and tilt parameters of PTZ camera. Li et al. [15] propose a mater-slave surveillance system for indoor scene. The method uses a mosaic image created by images of slave camera to estimate the relationship between static master camera image plane and pan-tilt controls of slave camera. All these systems mentioned above use static and active cameras. Bimbo et al. in Ref. [16] propose a novel framework exploiting two PTZ cameras to achieve the task of relating the feet position of a person in the image of the master camera to head position in the image of the salve camera. The system also uses a global map or a mosaic image to realize master-salve configuration. The main advantage of these systems mentioned above is that it could obtain global information and local multi-scale information. But all these works did not use depth information.

In this paper, a novel binocular vision system using two PTZ cameras for indoor scene is presented. Different from other systems, the proposed system can work on multimode: asymmetrical mode and symmetrical mode. In asymmetrical mode or master-salve mode, we propose a rectification-disparity-based method to establish correspondence between two image coordinates, and it does not need any pre-configuration, such as image mosaic, mapping construction, etc. Then, the salve camera could capture high resolution image to assist master camera' task. In symmetrical mode, we use images from the two cameras to estimate the depth of object located in common field of view (FOV). The calculated depth information is used to help solve the occlusion problem in tracking. The experimental results on indoor surveillance environments have demonstrated the effectiveness and robustness of our system. The paper is organized as follows: Section II gives an overall description of the PTZ camera calibration. Section III describes the stereo rectification method of dual PTZ cameras. Section IV introduces our proposed system, which consists of two modules, i.e. high resolution visual attention and object tracking using depth information. Section V gives the experimental results. Section VI summarizes the paper.

II. PTZ CAMERA CALIBRATION

Calibration of PTZ camera plays an important role in the visual system using PTZ cameras. In our system, PTZ camera calibration is an elementary preparation. This preparation can be done off-line. Our method is similar to [17] which is feature based combined with the parameters inquired from the PTZ camera. In order to calibration the PTZ camera, we first make some assumptions as follows:

(1) The rotation axes of pan and tilt are orthogonal and intersect at the origin of the camera coordinate system. In this cases, the extrinsic matrix can be directly calculated from the pan and tilt parameters.

(2) The camera's pixel aspect ratio $\alpha = 1$, and skew $s = 0$.

(3) The principal point (u_0, v_0) is fixed, and we use zoom center to replace the principal point.

With the rapid improvement of camera manufacturing, these assumptions can be well satisfied. Based on the above assumptions, the simplified camera model can be written as:

$$x = \kappa \boldsymbol{K}(Z)\boldsymbol{R}(P,T)\boldsymbol{X} \qquad (1)$$

where x and $\boldsymbol{X}$ are image coordinates and world coordinates respectively, κ is a scale factor. In (1), $\boldsymbol{K}(Z)$ is the intrinsic matrix which can be determined by the zoom parameter:

$$\boldsymbol{K}(Z) = \begin{bmatrix} f(Z) & 0 & u_0 \\ 0 & f(Z) & v_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (2)$$

The extrinsic matrix $\boldsymbol{R}(P,T)$ can be determined by the pan and tilt parameters which can be acquired from the camera:

$$\boldsymbol{R}(P,T) = \begin{bmatrix} \cos(P) & 0 & \sin(P) \\ -\sin(P)\sin(T) & \cos(T) & \cos(P)\sin(T) \\ -\sin(P)\cos(T) & -\sin(T) & \cos(P)\cos(T) \end{bmatrix} \qquad (3)$$

A. Principal Point Estimation

As mentioned above in assumption (3), we use zoom center to replace the principal point. To estimate the zoom center, we capture successive images $I_0, \cdots, I_z$ from the camera at varying zoom level with fixed pan and tilt parameters, where z is the maximum zoom level of the camera. Then we detect and match SIFT feature points [18] between $I_k (k = 1, ..., z)$ and I_0. For each of the matched points, we can obtain a straight line in the image plane. Ideally all the straight lines should converge at one point, i.e. the zoom center. So zoom center can be calculated by (4). Where $y = k_i x + b_i$ $(i = 1, \cdots, n)$ is the straight line determined by each of the matched points, n is the total number of the matched points.

$$\min_{x,y} \sum_{i=1}^{n} \frac{(y - k_i x - b_i)^2}{k_i^2 + 1} \qquad (4)$$

Taking derivatives with respect to x and y. Then we have

$$\begin{bmatrix} \sum_{i=1}^{n} \frac{k_i}{k_i^2+1} & \sum_{i=1}^{n} \frac{-1}{k_i^2+1} \\ \sum_{i=1}^{n} \frac{k_i^2}{k_i^2+1} & \sum_{i=1}^{n} \frac{-k_i}{k_i^2+1} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \frac{-b_i}{k_i^2+1} \\ \sum_{i=1}^{n} \frac{-k_i b_i}{k_i^2+1} \end{bmatrix} \qquad (5)$$

The solution of the above equation can be viewed as the estimation of the zoom center. Considering the existence of mismatching points, we use RANSAC method [19] to obtain a robust estimation of the zoom center finally. Fig. 1 shows our experiment result, where the black lines are the straight lines determined by the

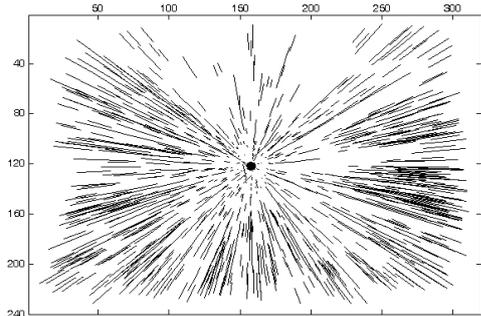matched points, and the solid circle is the estimated zoom center. The size of image is 320×240.

Figure 1.   Zoom center estimation result.

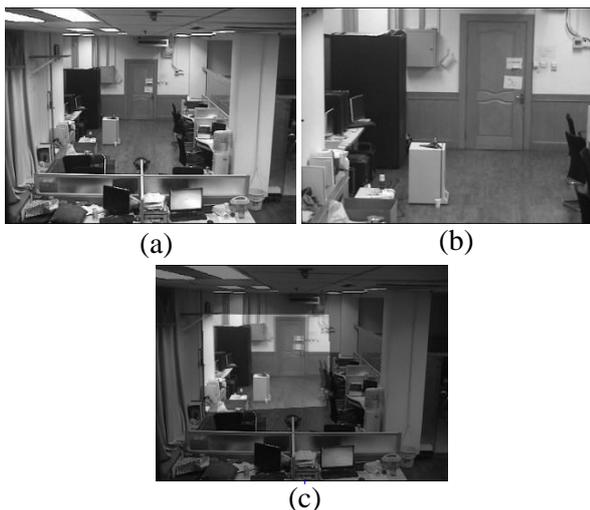(a)   (b)

(c)

Figure 2.   PTZ camera calibration verification. The top two images are the origin images captured from one of the two PTZ cameras. The bottom image is the warping result according to our calibration procedure. The PTZ parameters: (a) $PTZ_1 = [91.35, -11.23, 2.00]$ and

(b) $PTZ_2 = [87.44, -9.23, 9.00]$.

B. Focal Length Estimation at Different Zoom Level

As shown in (2), the focal length is determined by the zoom parameter. So we first estimate focal length at a specified zoom value Z. Two images under the same zoom value but different pan and tilt parameters are captured. We assume the parameters of the two images are (P, T, Z) and (P', T', Z) respectively. Then we use SIFT feature points to establish correspondence between the two images. For each of the matched points, we have

$$\begin{cases} \boldsymbol{x}_i = \kappa \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{R}(P, T) \boldsymbol{X}_i \\[3mm] \boldsymbol{x}_i' = \kappa \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{R}(P', T') \boldsymbol{X}_i \end{cases} \quad (6)$$

As the extrinsic matrix $\boldsymbol{R}$ can be calculated by (3), and the principal point (u_0, v_0) has been estimated in previous

content. So solving (6), we can obtain the estimation of focal length for each of matched points. In order to improve the robustness, we use all the matched points to calculate the mean value as the final result of focal length.

After focal length has been estimated at several discrete zoom levels, we choose a proper model to fit these samples. In our study, we use (7) for approximation. The four unknown parameters p_1, p_2, p_3 and p_4 can be solved by using curve fitting tools.

$$f(Z) = p_1 e^{p_2 Z} + p_3 e^{p_4 Z} \quad (7)$$

To verify the effectiveness of our calibration procedure, we capture two images from the PTZ camera shown in Fig. 2 (a) and (b) at different parameters. We use the camera calibration result to warp image (b) to image (a) with the same image coordinates. The result is shown in Fig. 2 (c). The result shows that the warped image fits the origin image well.

III.   STEREO RECTIFICATION USING DUAL PTZ CAMERAS

In traditional dual static system, the two cameras have the same intrinsic parameters then the acquired images are known as a rectified pair of stereo images. In these images, corresponding points lie on a same horizontal scan lines. So stereo matching algorithm can be directly applied to calculate the disparity or depth information. In dual PTZ camera system, it's hard to directly estimate disparity because that the two cameras may have different extrinsic and intrinsic parameters especially the focal length. Stereo rectification is a way to reduce the searching scope in stereo matching from 2D to 1D. Many stereo rectification methods have been proposed in the past years [20-24]. In this paper, we use spherical rectification method [23] to rectify images captured from PTZ cameras. The reason for choosing this method lies in that it can rectify images under arbitrary PTZ settings. Meanwhile, disparity obtained by this method could reflect the depth of the scene.

The main idea of spherical rectification method is using two spherical surfaces as the bridge to achieve fast and robust stereo rectification. It includes three steps: (1) estimate a rotation matrix and establish spherical coordinate system for each of the PTZ camera, the goal of this step is to make sure the longitude components of corresponding points be the same after coordinate conversion from the camera coordinate to spherical coordinate; (2) use PTZ camera model to map the image plane to the unit spherical surface according to current PTZ parameters; (3) the rectification is applied from the sphere to rectified plane. This method is independent to specific PTZ parameters, which is very convenient in application. More details can be found in [23]. Fig. 3 shows an example of spherical rectification for two images with quite different PTZ settings.

In our system, we use spherical rectification method to establish correspondence between two image coordinates. The conversion between origin image coordinate and rectified image coordinate are frequently used in our

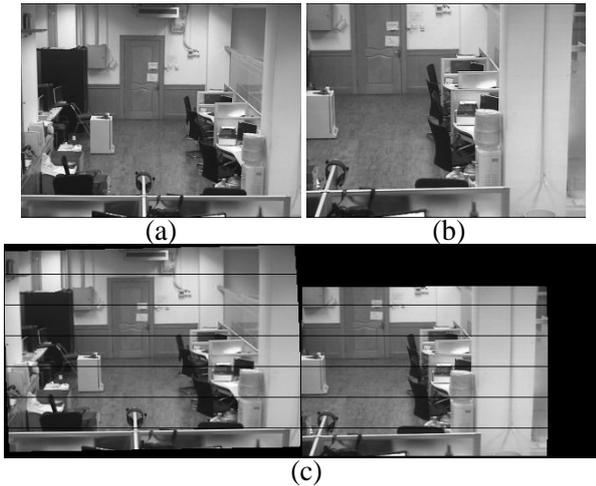system. So we summarize the conversion procedure as follows.

Figure 3. Example of spherical rectification: (a) and (b) are the origin images captured from the two PTZ cameras with different parameters; (c) is the rectification result. The PTZ parameters: (a) $PTZ_1 = [92.31, -10.65, 5.00]$ and (b) $PTZ_2 = [93.28, -12.62, 6.31]$.

C. From Origin Image Coordinate (x, y) to Rectified Image Coordinate (x_r, y_r)

Step1. Get camera coordinate X by the camera model $X = \kappa R^{-1} K^{-1} x$, where κ is a scale factor and $x = [x, y, 1]^T$. R and K are the extrinsic matrix and intrinsic matrix respectively, and can be obtained from the calibration result (see (2) and (3)).

Step2. Use (8) to transform X to spherical coordinate and obtain longitude and latitude coordinate (α, β). Where R_r is rotation matrix which can transform the camera coordinate to spherical coordinate. $X^r(m)$ is the m th component of vector X^r.

$$\begin{cases} X^r = R_r X \\ \alpha = \arctan(X^r(3), X^r(2)) \\ \beta = \arccos(X^r(1)) \end{cases} \quad (8)$$

Step3. Adjust (α, β) to (α, γ) by $\gamma = -\cot \beta$ so that the disparity can be preserved [23].

Step4. Calculate linear transformation parameters and transform (α, γ) to (x_r, y_r).

D. From Rectified Image Coordinate (x_r, y_r) to Origin Image Coordinate (x, y)

Step1. Linear transform (x_r, y_r) to (α, γ) by using the calculated transformation parameters.

Step2. Calculate the longitude and latitude coordinates (α, β) by $\beta = \tan^{-1}(-1/\gamma)$.

Step3. Get corresponding camera coordinate X from (α, β) by (9). Where R_r^{-1} is the inverse matrix of R_r.

$$\begin{cases} X^r(1) = \cos \beta \\ X^r(2) = \sin \beta \cos \alpha \\ X^r(3) = \sin \beta \sin \alpha \\ X = R_r^{-1} X^r \end{cases} \quad (9)$$

Step4. Get origin image coordinate (x, y) from the camera coordinate X by camera model (see (1)).

IV. THE PROPOSED SYSTEM

A. System Overview
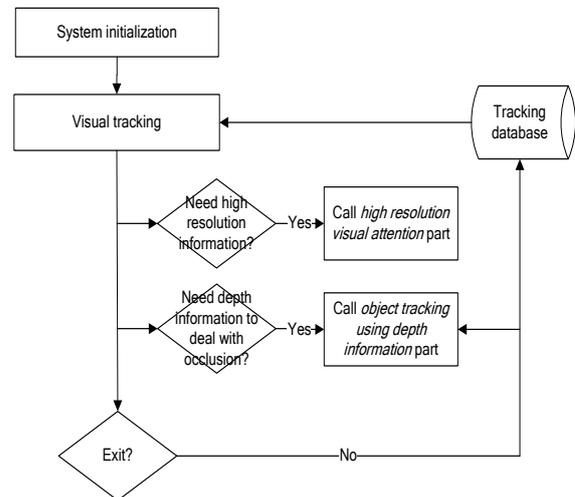
Figure 4. System flow chart.

For our system, the hardware configuration is composed of two PTZ cameras. The flow chart of our system is shown in Fig. 4. Before main procedure starts, an initialization procedure is needed. We first choose one camera as the static view, and the other camera will be served as the active camera. Then, in order to reduce the computational complexity, some parameters can be pre-calculated, i.e. camera calibration, stereo rectification parameters, etc. The main procedure is composed of one main thread and some other assistant threads. The main thread is a common visual tracking loop in the static view, including foreground extraction, objects detection and tracking. The other assistant threads include high resolution visual attention and object tracking using depth information. Both of the two assistant threads need the active camera's cooperation. The visual tracking in static view has been well studied in past decades. So we mainly consider the two parts, i.e. the high resolution visual attention and object tracking using depth information. In high resolution visual attention part, two cameras work on asymmetrical or master-salve module, and we use the stereo rectification result and disparity to build the relationship between two camera image coordinates. Therefore, visual attention of multi-resolution can be obtained. In object tracking using depth information part, two cameras work on symmetrical mode, and we estimate depth information on the rectified image fair. So that it can be used for object tracking under the situation of occlusion. In summary, the proposed system has the

advantage of both traditional dual static and static-active system.

B. High Resolution Visual Attention

In high resolution visual attention part, the two cameras work on asymmetrical mode or master-salve mode. The key technique of high resolution visual attention is to calculate the new PTZ parameters of the salve camera according to the master's PTZ setting and the selected region in master camera's image. So the slave camera could cover that region with a higher resolution. Traditional methods mainly have two kinds of control strategies: sampling-interpolation method and mosaic image-based method. The first method constructs a mapping from every pixel coordinates in static image to PT parameters (such as [12]). This method has two drawbacks. First, it assumes that the baseline is too small comparing to the depth of the scene. So if the scene has great change in depth, such as indoor environment, this method may have large error. Second, if the master camera's view changes, the pre-configurations should be re-done. The second method uses a mosaic image created by snapshots of slave camera to estimate the relationship between static master camera image plane and pan-tilt controls of slave camera (such as [15, 16]). This method relies on feature matches of master camera image and salve camera mosaic image, and when the scene appearance changes, the mosaic image should be updated.

To overcome the drawbacks of these methods, we proposed another approach which directly uses rectified image and disparity to calculate new PTZ parameters. Rectified image pair and disparity map is a bridge between the two images coordinates, so that two image coordinates can be mapped from one to another by rectification mapping added a disparity translation. We denote I_1 and I_2 as the master camera image and salve camera image respectively, (P_1, T_1, Z_1) and (P_2, T_2, Z_2) as the parameters of the master and salve camera. We denote $\Re$ as the selected region in master camera's image. The procedure of our method is listed as follows:

Step1. Use spherical rectification method to rectify I_1 and I_2. We denote the rectified images as I_{1r} and I_{2r}.

Step2. Find the center c_1 of the region $\Re$ in image I_1. Then calculate c_1 's corresponding location c_{1r} in rectified image I_{1r} (see detailed procedure in section III).

Step3. Use graph cuts based stereo matching algorithm [25] to estimate the mean disparity d at c_{1r}. Denote c_{2r} as the corresponding point in rectified image I_{2r} of the salve camera image I_2, where $c_{2r} = c_{1r} + [d, 0]^{\mathrm{T}}$.

Step4. Calculate c_2 in I_2 from corresponding c_{2r} in I_{2r} (see detailed procedure in section III).

Step5. Calculate new PTZ parameters (P_2', T_2', Z_2') given point c_2 and the origin PTZ parameters (P_2, T_2, Z_2) of the slave camera. This step can be divided into two parts, the PT parameters and Z parameter calculation. Let $c_2 = (x_s, y_s)$, and the principal point

is (u_0, v_0). According to (7) and zoom value of slave camera Z_2, we can calculate the focal length f. Then the absolute angles of pan and tilt parameters are

$$\begin{cases} \Delta P = \arctan \dfrac{\Delta x_s}{f} = \arctan \dfrac{|x_s - u_0|}{f} \\ \Delta T = \arctan \dfrac{\Delta y_s}{f} = \arctan \dfrac{|y_s - v_0|}{f} \end{cases} \quad (10)$$

Based on the position of point c_2, we use (11) to compute the new pan and tilt parameters, P_2' and T_2'. Fig. 5 is the sketch map for pan and tilt parameters estimation of slave camera.

$$\begin{cases} P_2' = P_2 + \Delta P, T_2' = T_2 - \Delta T & x_s \geq u_0 \ and \ y_s \geq v_0 \\ P_2' = P_2 + \Delta P, T_2' = T_2 + \Delta T & x_s \geq u_0 \ and \ y_s < v_0 \\ P_2' = P_2 - \Delta P, T_2' = T_2 - \Delta T & x_s < u_0 \ and \ y_s \geq v_0 \\ P_2' = P_2 - \Delta P, T_2' = T_2 + \Delta T & x_s < u_0 \ and \ y_s < v_0 \end{cases} \quad (11)$$
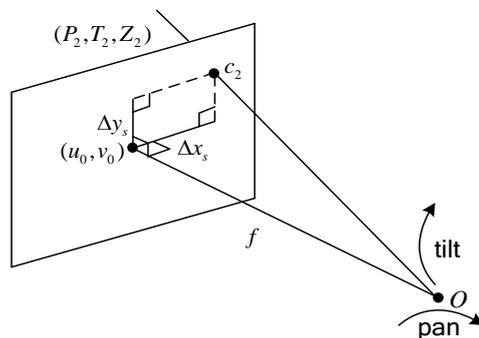
Figure 5. Sketch map for pan and tilt parameters estimation of slave camera.
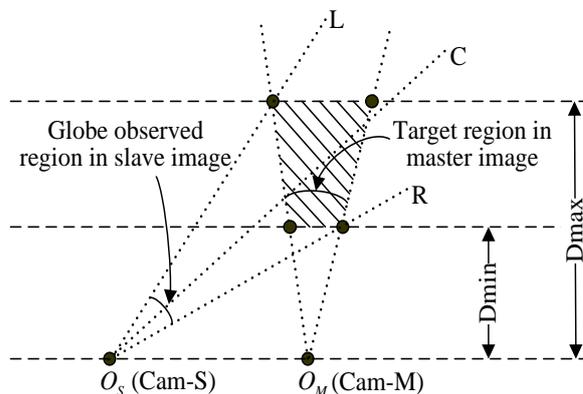
Figure 6. Sketch map for rough PTZ estimation by the maximum and minimum depth. Where $O_S C$ decides pan and tilt values, $O_S L$ and $O_S R$ decide the zoom value.

The new zoom parameter Z_2' of slave camera can be determined by the size of the bounding box of the given region. Firstly, we use a lookup table to get a zoom level Z_0 according to the size of given region bounding box. Then we define the reliability of disparity Y_d as a function of the variance of the local disparity in the target region,

which satisfies $Y_d \in [0,1]$. We use Y_d as a weight, so the final zoom parameter $Z_2^{'} = Z(Z_0, R_d)$ (in our system, $Z_2^{'} = Z_0(0.7 + 0.3R_d)$). That's means if the reliability of disparity is low, we lower down the zoom level. Otherwise, we give a higher zoom level.

If the target region is not visible in the salve camera image, we can't estimate the disparity translation on the two rectified images. In this situation, we first estimate rough PTZ parameters by the maximum $D_{\max}$ and minimum $D_{\min}$ depth, see Fig. 6. Then after the salve camera moves to the rough PTZ position, we use procedure described above to estimate precise PTZ parameters.

C. Object Tracking Using Depth Information

Depth information is very useful in visual surveillance. And the depth information can improve object segmentation and tracking in case of multiple occluding objects. All work before use dual static cameras. In this paper, the depth information is introduced to our dual PTZ camera system. In this part, the two cameras work on symmetrical mode. As we know, when occlusion happens, part of the occluded object is invisible. So the occluding and occluded objects have different depth. Although the depth information estimated by vision approach is always less credible because of the uncertainty of stereo matching, in our system, this coarse depth information is adequate in that we only need to give a depth order of occluded objects.

In our system, we implement a tracking loop in static camera procedure which works well with no occlusion. Under occlusion situation, the tracking procedure will send a message to active camera. Then two cameras can calculate depth information by spherical rectification and stereo matching. In order to reduce computational work, we only estimate the depth information in foreground area. The prior knowledge of segmentation and tracking comes from the tracking history which can be obtained from the tracking database, see Fig. 4. The prior knowledge includes the number of objects in the occlusion area and previous location of each object. Then we classify each foreground pixel into objects according to location and depth information of each pixel. To improve the robustness of our system, we also use a post process of validation. The validation considers the proportion of segmentation region. If the region is too small, the segmentation is invalid. For valid segmentation, the timestamp, segmentation region and depth information will be updated into tracking database.

We summarize the procedure as follows:

Step1. Use spherical rectification method to rectify images of the two cameras.

Step2. Estimate the disparity map in the foreground region by using graph cuts based stereo matching algorithm [25].

Step3. Collect prior knowledge from the tracking database, such as the objects number in occlusion area n and the previous location of each object.

Step4. Classify all the foreground pixels into n classes according its disparity and coordinates.

Step5. Calculate the property of each new class, such as area, centre and depth.

Step6. Validate the segmentation result. We consider the proportion of segmentation area, that a too small value suggests an invalid segmentation.

Step7. Update new class centre and depth into tracking database for valid segmentation result.

V.   EXPERIMENTAL RESULTS

The validity of our system has been evaluated on an indoor environment. The system runs on one computer with Intel 3.0G CPU and 1.5G memory. Two SONY EVI D70 cameras were placed on the top window, at about 1.2 meters far from each other. Images from both cameras were taken at 320×240 pixels of resolution.

(a)

(b)
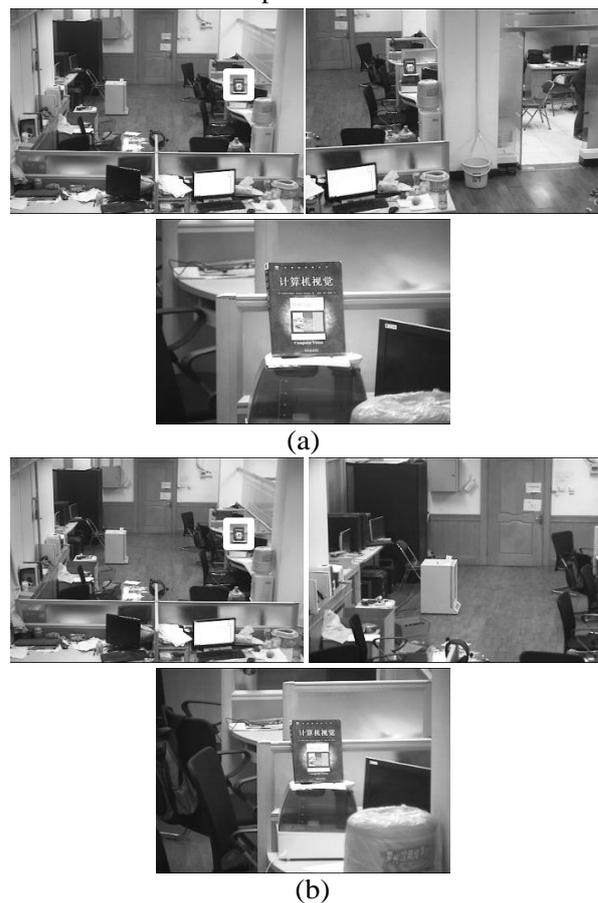
Figure 7.   Results of high resolution visual attention. The top two images of (a) and (b) are the origin images of the master and salve camera respectively. The bottom images of (a) and (b) are salve camera images after moving to newly estimated values. (a) The selected object is visible in the salve camera image; (b) The selected object is invisible in the salve camera image.

Firstly, we present the experimental results of high resolution visual attention. The experiments try to explain the robust result in two cases of salve camera's state: the selected object is visible in the salve camera image and the selected object is invisible in the salve camera image. Fig. 7 shows the results under these two conditions. In the initialization, the target region is selected in master camera image highlighted by white box. The parameters
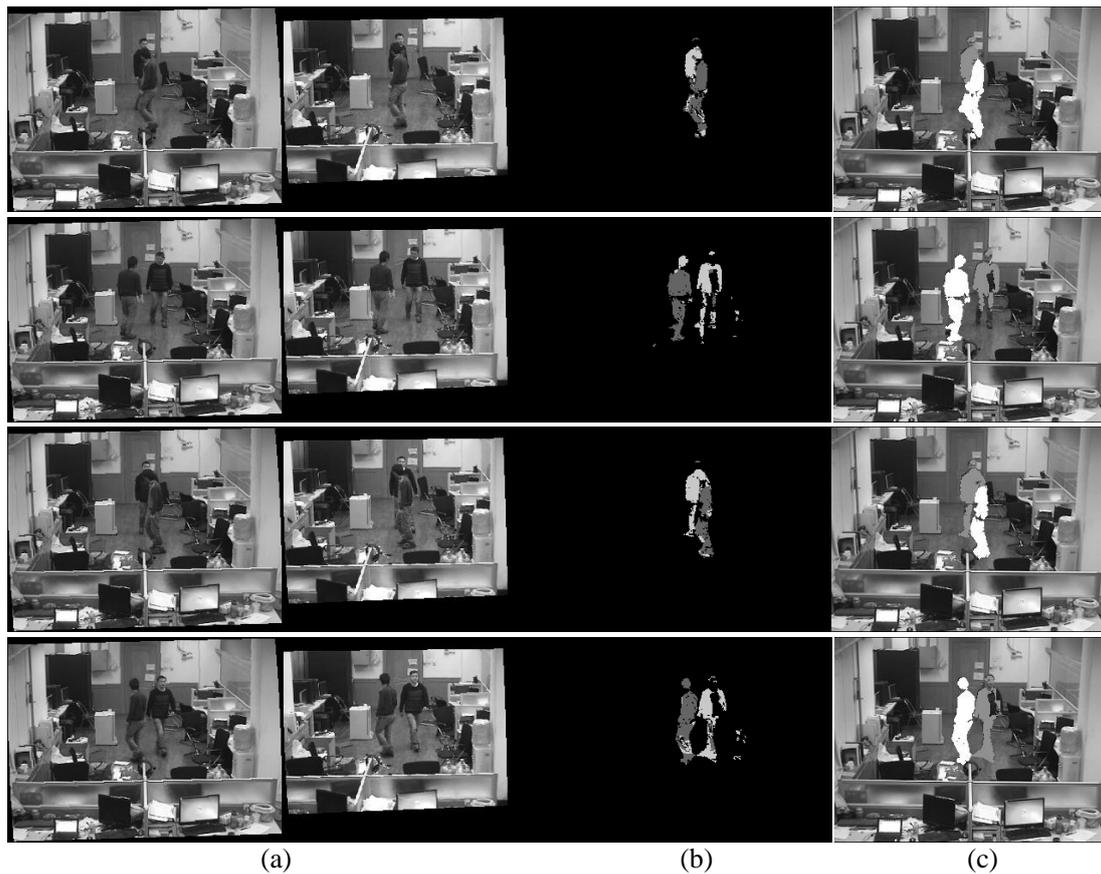
Figure 8.   Experimental results of object tracking under occlusion. (a) Rectified image pair; (b) Disparity map of foreground region; (c) Tracking result. The parameters of the two cameras are $PTZ_1 = [90.51, -14.55, 4.03]$ and $PTZ_2 = [86.91, -13.73, 5.65]$ respectively.

of master camera are $PTZ_1 = [90.44, -14.55, 4.03]$. In Fig. 7 (a), the initial parameters of the salve camera are $PTZ_2 = [99.14, -15.68, 5.11]$, and the selected object is visible in the salve camera image. We calculate the new PTZ parameters of the salve camera using method illustrated in section IV. The estimated parameters of salve camera are $PTZ_2^{'} = [94.93, -11.18, 14.33]$. In Fig. 7 (b), the initial parameters of salve camera are $PTZ_2 = [83.84, -10.88, 7.51]$, and the selected object is invisible in the salve camera's image. So we first use the given rough depth range $[2m, 15m]$ to calculate the rough parameters of salve camera. After the camera moves to the new setting, we estimate the precise parameters finally. The estimated parameters of salve camera are $PTZ_2^{'} = [94.71, -11.48, 12.81]$.

When occlusion happens in the static camera view, we use two cameras to estimate depth information, and the depth is used in object tracking in our system. The designed algorithm is based on disparity map, and no appearance features are used. We take the previous location as the prediction for current frame, and the goal is to find the new center of each person in current frame. Fig. 8 gives four groups of results. For each group, the left two images are rectified images captured from two cameras at the same time stamp. The third image is the disparity map of the foreground region, the greater the

intensity, the smaller the disparity, the greater the depth. The right image shows the tracking result. The two objects are presented by different intensity. The segmented region will be used to update the property of tracked object in tracking database. Note that the gray levels in different images are not comparable. This experiment is used to validate the depth information acquired from two PTZ cameras can be used in object tracking, especially when occlusion happens. And this application could handle those partly occlusion situation with not too small difference in depth.

VI.   CONCLUSION

In this paper, we propose a novel binocular vision system for indoor scene visual surveillance. Our system can work on multimode: asymmetrical mode (master-salve mode) and symmetrical mode. In asymmetrical mode, we propose a rectification-disparity-based method to establish correspondence between two image coordinates. Then, both global and detailed image information can be obtained. In symmetrical mode, we use depth information to deal with occlusion problem in object tracking. The proposed system has the advantages of both dual-static system and static-active or dual active system. The experimental results show the merits of the proposed system. In our future work, we plan to combine multiple other features and depth information to deal with occlusion problem in object tracking.

REFERENCES

[1] I. Ahmad, Z. He, M. Liao, F. Pereira, and M.-T. Sun, "Special issue on video surveillance", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, pp. 1001-1005, 2008.

[2] T. Gao, P. Wang, C Wang, and Z. Yao, "Feature particles tracking for moving objects", *Journal of Multimedia*, vol. 7, no.6, pp. 408-414, 2012.

[3] D. Kieran, J. Weir, and W. Yan, "A framework for an event driven video surveillance system", *Journal of Multimedia*, vol. 6, no.1, pp. 3-13, 2012.

[4] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: real-time tracking of the human body", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 780-785, 1997.

[5] C. Pai, H. Tyan, Y. Liang, H. Liao and S. Chen, "Pedestrian detection and tracking at crossroads", *Pattern Recognition,* vol. 37, pp.1025-1034, 2004.

[6] I. Haritaoglu, D. Harwood, and L.S. Davis, "W4: real-time surveillance of people and their activities", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 809-830, 2000.

[7] X Zhu, Y Gui, and H Xu, "Enhanced differentiated surveillance for static and random mobile sensor networks", *Journal of Networks*, vol. 6, no 10, pp. 1483-1490, 2011.

[8] G. Adorni, S. Cagnoni, M. Mordonini, and A. Sgorbissa, "Omnidirectional stereo systems for robot navigation", *in Proc. IEEE Workshop on Omnidirectional Vision and Camera Networks*, pp.79-89, 2003.

[9] C. Chen, Y. Yao, D. Page, B. Abidi, A. Koschan, and M.Abidi, "Heterogeneous fusion of omnidirectional and PTZ cameras for multiple object tracking", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no.8, pp. 1052-1063, 2008.

[10] R. Munoz-Salinas, E. Aguirre, M. Garcia-Silvestre, and A. Gonzalez, "A multiple object tracking approach that combines colour and depth information using a confidence measure", *Pattern Recognition Letters*, vol. 29, pp. 1504-1514, 2008.

[11] R. Munoz-Salinas, "A Bayesian plan-view map based approach for multiple-person detection and tracking", *Pattern Recognition*, vol. 41, pp. 3665-3676, 2008.

[12] X.H. Zhou, R.T. Collins, T. Kanade, and P. Metes, "A master-slave system to acquire biometric imagery of humans at a distance", *ACM SIGMM International Workshop on Video Surveillance*, pp. 113-120, 2003.

[13] A.W. Senior, A. Hampapur, and M. Lu, "Acquiring multi-scale images by pan-tilt-zoom control and automatic multi-camera calibration", *IEEE Workshop on Application on Computer Vision*, pp. 433-438, 2005.

[14] R. Bodor, R. Morlok, and N. Papanikolopoulos, "Dual-camera system for multi-level activity recognition," *Proc. of the IEEE/RJS International Conference on Intelligent Robots and Systems*, vol. 1, pp. 643-648, 2004.

[15] Y. Li, L. Song, and J. Wang, "Automatic weak calibration of master-salve surveillance system based on mosaic image", *International Conference on Pattern Recognition*, pp. 1824-1827, 2010.

[16] A.D. Bimbo, F. Dini, G. Lisanti, and F. Pernici, "Exploiting distinctive visual landmark maps in pan-tilt-zoom camera networks", *Computer Vision and Image Understanding*, vol. 114, pp. 611-623, 2010.

[17] S.N. Sinha and M. Pollefeys, "Pan-tilt-zoom camera calibration and high-resolution mosaic generation", *Computer Vision and Image Understanding*, vol. 103, pp. 170-183, 2006.

[18] D.G. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60, pp. 91-101, 2004.

[19] M.A. Fischler, R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography", *Communications of ACM*, vol. 24, pp. 381-395, 1981.

[20] M. Pollefeys and S.N. Sinha, "Iso-disparity surfaces for general stereo configurations", *European Conference on Computer Vision*, vol. 3, pp.509-520, 2004.

[21] C.T. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision", *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1125-1131, 1999.

[22] R.I. Hartley, "Theory and practice projective rectification", *International Journal of Computer Vision*, vol. 35, pp. 115-127, 1999.

[23] D. Wan, J. Zhou, and D. Zhang, "A spherical rectification for dual-PTZ camera system", *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 777-780, 2007.

[24] S. Kumar, C. Micheloni, C. Piciarelli, and G.L. Foresti, "Stereo rectification of uncalibrated and heterogeneous images", *Pattern Recognition Letters*, vol. 31, pp. 1445-1452, 2010.

[25] Y. Boykov, O. Veksler, and R. Zabin, "Fast approximate energy minimization via graph cuts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1222-1239, 2001.

Zhigao Cui born in 1984, Hengshui, Hebei province, China. He is currently a doctor of Xi'an Institute of High-tech of China. His research interests include binocular vision surveillance and image processing.

Aihua Li was born in 1966. Professor and Ph.D. supervisor in Xi'an Institute of High-tech of China. His main research interests include pattern recognition, signal processing and computer vision.

Classification-based Multi-client Video Transmission over Heterogeneous Networks

Bo Li, Peng Wang, and Yongfei Zhang

Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China

Email: bhboli@vip.sina.com, knightwp@126.com, zhangyf.ac@gmail.com

Abstract—These Real-time video streaming over networks operates under stringent network resource constraints, with multiple video clients competing for limited network resources. In this paper, we study the problem of bandwidth allocation for video transmission over heterogeneous networks, with multiple video clients connecting to the video server simultaneously and demanding for the video services, and aim to provide the best possible Quality of Service (QoS) under limited bandwidth of both the video server and multiple video clients. We propose a classification-based approach for multi-client video transmission over heterogeneous networks (CMVT). Firstly, the video server detects the available bandwidth of multiple video clients and classifies the clients into different classes. Secondly, the limited export bandwidth of the server is allocated to different video clients using the client classification results and greedy algorithm. Finally, the video server transmits video streams to video clients in different classes through Unicast and clients in the same class through Unicast and forwarding. Experimental results demonstrate that the proposed video transmission method can use the network bandwidth efficiently and provide better video quality to more video clients.

Index Terms—Video Transmission, Client Classification, Bandwidth Allocation, Unicast and Forwarding

I. INTRODUCTION

With the fast development of the real-time video demands and networking techniques, real-time video applications are playing more and more important role in the industrial area as well as our daily lives, a lot of video transmission schemes have also been proposed in the literature.

Generally, in real-time video transmission, a single video stream cannot satisfy the requirements of multiple video clients in a heterogeneous network environment [1], because the available bandwidth of each video client might be different. What is more, when multiple video clients connect to the video server simultaneously, they have to compete for the limited export bandwidth of the video server. If the limited export bandwidth cannot be allocated to video clients reasonably, the reconstructed video quality provided to the video clients would severely degrade.

The most commonly used traditional transmission methods are Unicast, Multicast and Broadcast. Unicast is a commonly used transmission scheme which establishes one connection between each sender and receiver. The control of Unicast is flexible, but it cannot efficiently utilize the network bandwidth. Multicast is used to transmit data from a single sender to multiple receivers simultaneously which can efficiently utilize the network bandwidth. However, the flexibility and security of Multicast is poor [2], [3]. The receivers may attempt to request different video quality with different bandwidth, but only one video stream is sent out from the video server when using Multicast, which cannot meet various requirements of the receivers. Broadcast transmits data from one sender to all the receivers, regardless of whether the receivers need or not. Several Broadcast schemes [4] such as fast broadcast, harmonic broadcast and staggered broadcast have been proposed, and it has been proved that the harmonic broadcast is more efficient than the other broadcast schemes [5]. In [6], [7], video transmission methods combining Broadcast and Unicast are proposed. These methods may consume minimal export bandwidth of the video server, but they may also introduce the broadcast storms easily, which will results in severe network congestions and significant performance degradation. What is more, similar to Multicast, the safety and flexibility issues prevent Broadcast from many practical video applications. Considering the transmission safety and data privacy, most network intermediate devices such as routers/switches in current networks like Internet do not support Multicast and Broadcast and only support Unicast.

In most practical applications, the export bandwidth of the video server is limited, and the resulting transmission bottleneck on the video server side cannot be avoided [8]. So how to allocate the limited export bandwidth and provide better video quality to satisfy the requirements of different video clients becomes a critical issue. Several bandwidth allocation algorithms have been proposed in [9], [10]. These algorithms are either operated on the network intermediate devices or not aiming at the characteristics of the video transmission. But in most practical applications it is not allowed to access the intermediate network devices. More practical solutions are those end-system based schemes which do not require the participation of the networks and are applicable to both the current and future Internet. In [11]-[15], several

rate-distortion optimization-based video transmission algorithms have been proposed. They improve the reconstructed video quality considering the rate-distortion characteristics of the video. But these algorithms are proposed only for single video transmission. When multiple video streams are transmitted in the same network link, the quality of all the video streams should be considered jointly. Moreover, some of these algorithms may bring long capture-to-display latency (CDL) [16] which can not satisfy the real-time requirements in the video applications, especially for the video applications with interactions between the video clients and the video server. In [17], an efficient bandwidth resource allocation algorithm, M-LDLF (Multiuser Low-delay resource allocation algorithm with Low-fluctuation), is proposed. M-LDLF provides low fluctuation of quality for each user and consistent quality among all users by exploring the variation in the scene complexity of each video program and jointly redistributing available system resources among users. Given the limited export bandwidth of video server, M-LDLF may achieve good performance with small number of clients connect to the video server simultaneously. But if there are more clients, the transmission scheme adopted in M-LDLF that one encoder corresponds to one user would cause severe degradation in the reconstructed video quality.

In real-time video applications, when multiple clients with different available bandwidth connect to the video server simultaneously, only one video stream cannot satisfy all the requirements of these clients. On the other hand, it is too expensive or even impossible to generate and transmit a video stream with specific bitrate to accommodate each client using Unicast with guaranteed video quality, since the computational and storage resource as well as the export bandwidth of the video server are limited.

In this paper, we alleviate this dilemma with a novel client-classification-based approach. Aiming to enable more video clients to access the video server simultaneously and obtain better reconstructed video quality under limited export bandwidth of the video server, we propose a classification-based multi-client video transmission (CMVT) scheme over Heterogeneous networks. Firstly, we classify video clients into different classes according to their available bandwidth. Secondly, the greedy algorithm is adopted to allocate the export bandwidth of the video server to different classes using client classification results. Finally, a hybrid video transmission scheme is developed, in which Unicast is used to transmit video streams to the clients in different classes, Unicast and forwarding are both used to transmit video streams to the clients in the same class. The rest of the paper is organized as follows. Section II describes the discriminant analysis-based client classification scheme. In Section III, we present the proposed greedy and client classification-based bandwidth allocation algorithm. The hybrid video transmission scheme is developed in Section IV, followed by the proposed classification-based multi-client video transmission scheme. Section V presents the

experimental results and performance comparisons. Finally, the paper is concluded in section VI.

II. DISCRIMINANT ANALYSIS-BASED CLIENT CLASSIFICATION

In this section, we present the discriminant analysis-based client classification scheme. When a video client connects to the video server, its available bandwidth can be detected using bandwidth detection method firstly, like that in [18].

The client classification is a typical problem of discriminant analysis. Since the goal of the proposed classification is to enable more video clients to access the video server simultaneously and improve the reconstructed video quality that the video clients receive under limited export bandwidth, the discriminant rule should satisfy the following two conditions: 1) the video clients with close available bandwidth should be classified into the same class and, 2) the bandwidth center of each class should have sufficient difference to each other. We choose the distance discriminant analysis [19] as the discriminant analysis methods and the Mahalanobis distance (MD) [20] as the discriminant rule to solve the video client classification problem. Note that the abovementioned two conditions can be appropriately described by the covariance and mean value of the bandwidth in one class. The covariance and mean value represent the degree of dispersion and the center of each class, respectively.

In statistics, the MD is based on the correlations between variables and measures the similarity of an unknown sample set to a known one. It outperforms the Euclidean distance by taking into consideration of the correlation of the data set and is scale-invariant. The MD of x to the class C_i can be calculated as follows.

$$D(x, C_i) = \sqrt{\frac{(x - \mu_i)^2}{\delta_i}} \qquad (1)$$

where x is the new sample, which is the available bandwidth of the new joined client in our scheme. μ_i and δ_i are the mean value and covariance of the samples in the class C_i, respectively. In our client classification scheme, they represent the center and the degree of dispersion of the bandwidth of different video clients in one class. $D(x, C_i)$ is the Mahalanobis distance from x to C_i. The smaller the value of MD is, the higher the probability of x belongs to C_i.

In this paper, the discriminant function is formulated as:

$$W_{ij} = D^2(x, C_i) - D^2(x, C_j) = \frac{(x - \mu_i)^2}{\delta_i} - \frac{(x - \mu_j)^2}{\delta_j} \qquad (2)$$

where W_{ij} represents the discriminant result, and $\delta_i \times \delta_j \neq 0$. If $i \neq j$ and $W_{ij} < 0$, then $x \in C_i$; otherwise, if $i \neq j$ and $W_{ij} > 0$, then $x \in C_j$; if $W_{ij} = 0$, then x belongs to the class whose δ is bigger than the others,

because δ indicates the dispersion degree of a class. The bigger δ is, the higher the dispersion degree is. If the new client x joins the class with bigger dispersion degree, the impact on the dispersion degree would be smaller. Otherwise the impact may be bigger. As shown in Fig. 1, the dispersion degree of Class J (denoted by squares) is bigger than that of Class I (denoted by circles). When a new client x (denoted by red dot) joins, if $W_{ij} = 0$, x would be classified into class J.

If $\delta_i \times \delta_j = 0$, we use (3) to classify clients,

$$W_{ij} = (x - \mu_i)^2 - (x - \mu_j)^2 \qquad (3)$$

In (3), if $i \neq j$ and $W_{ij} < 0$, then $x \in C_i$; otherwise, if $i \neq j$ and $W_{ij} > 0$, then $x \in C_j$; if $W_{ij} = 0$, x belongs to the class whose $\delta = 0$.
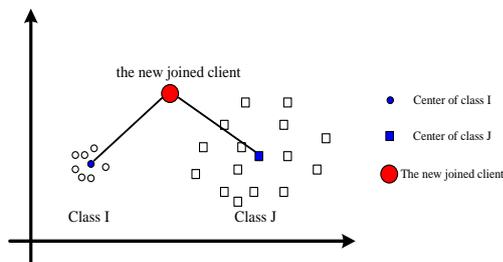

Figure 1.   Client classification.

So far, we can classify the video clients into different classes according to their available bandwidth using the distance discriminant analysis. The problem now becomes: How many classes the video clients should be classified? The more classes the video clients are classified into, the more the requirements of different clients can be satisfied. But more classes also mean that more export bandwidth and more computational and storage resources of the video server are required. If the resources of the video server are not enough, they would become the bottlenecks of video transmission. On the contrary, if the video clients are classified into fewer classes, the export bandwidth and the other resources would be no longer the bottleneck, but the requirements of different video clients would not be fully satisfied.

Typically, the current network transmission links can be classified into three types according to the available bandwidth. The first type includes the links of broadband networks or those constructed or leased specially for video transmission, where the available bandwidth of more than 1Mbps can be usually guaranteed. The second type is that the links in public networks or some private networks are used to transmit many kinds of data, where the available bandwidth is usually around 200Kbps~600Kbps, such as the uplinks of ADSL. The third type is those low bandwidth links, like 2G/2.5G/3G wireless links, where the available bandwidth is usually around 30Kbps~150Kbps.

Considering the abovementioned practical heterogeneous network infrastructures, in this paper, we classify the video clients into three classes: the high-

bandwidth class, the mid-bandwidth class and the low-bandwidth class. Corresponding to the three classes, the video server starts three video encoders: the high-bandwidth encoder, the mid-bandwidth encoder and the low-bandwidth encoder. If a video client connecting to the video server is classified into the high-bandwidth class, the high-bandwidth encoder would start to work to output higher-resolution (720×576) video streams to the video client. The mid-bandwidth encoder outputs lower-resolution (352×288) video streams to the video clients in the mid-bandwidth class. If the available bandwidth of a video client is relatively poor, the low-bandwidth encoder will be triggered to output mid-resolution but low-frame rate video streams to satisfy the requirements of the video clients.

III.   DYNAMIC BANDWIDTH ALLOCATION BASED ON GREEDY ALGORITHM AND CLIENT CLASSIFICATION

In the real-time video transmission, the reconstructed video quality provided to the clients need to be guaranteed [21], [22]. However, the available bandwidth of each client is usually different and limited under the complex and heterogeneous network infrastructures. Although in Section II, we use (2) and (3) to classify clients into three classes according to their available bandwidth, the reconstructed video quality would severely degrade if the limited export bandwidth of the video server cannot be allocated reasonably.

In [1], by analyzing the real-time video transmission techniques, two general approaches, network-centric approaches and end system-based approaches have been proposed. The network intermediate devices such as routers/switches are required to provide QoS support to guarantee bandwidth, bounded delay and packet loss for video applications for the network-centric approaches. But in most practical applications, we cannot access the network intermediate devices and impose any requirements on the network. So in this section, we propose an end system-based approach called dynamic bandwidth allocation based on greedy algorithm and client classification to maximize the overall reconstructed video quality to all video clients. This bandwidth allocation algorithm operates on the video server and does not impose any requirements on the network. Given the export bandwidth of the video server, we can allocate the export bandwidth to different classes reasonably, especially when the export bandwidth is limited. The encoders in the video server can adaptively adjust their output bitrates accordingly and send them to different video clients. This bandwidth allocation algorithm can efficiently utilize the limited bandwidth, thus improve the reconstructed video quality provided to the video clients.

In this paper, when some bandwidth is allocated to a class, the corresponding encoder will adjust its output bitrate to adapt to the allocated bandwidth, so we can regard the allocated bandwidth to a class as the output bitrate of the corresponding encoder. The Rate-Distortion (R-D) theory [23] gives the relationship between the output bitrate of the encoder and the distortion of the reconstructed video, as shown in Fig. 2.
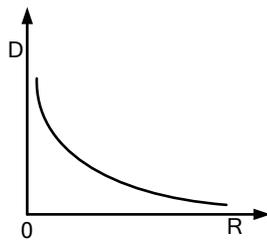
Figure 2.   The R-D curve.

It also can be seen from Fig. 2 that the higher the bandwidth is allocated to a sequence, the smaller the distortion of the reconstructed video is, and vice versa. What is more, the less the bandwidth is, the bigger the improvement of reconstructed video quality is when the same amount of additional bandwidth is allocated.

In order to make the best use of the limited network bandwidth and provide better video quality to the video clients, we propose a bandwidth allocation algorithm based on greedy algorithm and client classification. The greedy algorithm is adopted to determine how many video clients can connect to the video server directly and how much bandwidth should be allocated to each class. The principle of the greedy bandwidth allocation is to enable higher bandwidth clients to connect to the video server directly, because the reconstructed video quality provided to the clients with higher bandwidth is better and the capture-to-display latency of direct-connected clients is smaller. During the video transmission, if there is still remaining bandwidth and the transmitted bitrates are lower than the available bandwidth of the classes, the classes would gain some remaining bandwidth to improve its received video quality. The principle of the remaining bandwidth allocation is to improve the reconstructed video quality as much as possible with the same amount of additional bandwidth. The proposed dynamic bandwidth allocation algorithm can be described as follows.

(1) Initialization.

Set the initial bandwidth center of each class for client classification at the beginning, denoted as B_h^{ini}, B_m^{ini} and B_l^{ini}, corresponding to the three types of heterogeneous transmission links. In this paper, we set that $B_h^{ini} = 1300 Kbps$, $B_m^{ini} = 300 Kbps$ and $B_l^{ini} = 50 Kbps$. With more clients joining in and being classified into the classes, the bandwidth center can be updated as the average available bandwidth of clients in the class.

(2) Rough bandwidth allocation.

Determine the bandwidth that should be allocated to the high-bandwidth class, the mid-bandwidth class and the low-bandwidth class, denoted as B_h, B_m and B_l by the following formulas, respectively.

$$B_h = \min_k\{B_h^k, B_h^{ini}\}$$
$$B_m = \min_k\{B_m^k, B_m^{ini}\} \qquad (4)$$
$$B_l = \min_k\{B_l^k, B_l^{ini}\}$$

where B_h^k, B_m^k and B_l^k are the available bandwidth of the k-th client in the high-bandwidth class, the mid-bandwidth class and the low-bandwidth class, respectively. We bound the allocated bandwidth to each class with the initial bandwidth, as in (4), to avoid that too much bandwidth might be allocated to the high bandwidth class and the clients in the low bandwidth class might get starved at the very beginning of the bandwidth allocation. Then the corresponding encoders adjust their output bitrates to adapt to the allocated bandwidth.

Determine the number of direct-connected video clients in the high-bandwidth class, $n_h = \min\{\frac{B_{total}}{B_h}, N_h, N_{total-h}\}$, where N_h denotes the maximal number of video clients in the high-bandwidth class that the video server allows to connect it directly, $N_{total-h}$ denotes the total number of video clients in the high-bandwidth class. B_{total} denotes the total export bandwidth of the video server. Then the number of direct-connected video clients in the mid-bandwidth class and low-bandwidth class can be determined as follows.

$$n_m = \min\{\frac{B_{total} - n_h B_h}{B_m}, N_m, N_{total-m}\} \qquad (5)$$

$$n_l = \min\{\frac{B_{total} - n_h B_h - n_m B_m}{B_l}, N_l, N_{total-l}\} \qquad (6)$$

where N_m and N_l denotes the maximal number of video clients in the mid-bandwidth class and low-bandwidth class that the video server allows to connect it directly, $N_{total-m}$ and $N_{total-l}$ denotes the total number of clients in the mid-bandwidth class and low-bandwidth class, respectively.

(3) Fine bandwidth allocation.

If the total bandwidth allocated to the direct-connected video clients is less than the export bandwidth of the video server, and the allocated bandwidth is lower than the available bandwidth of any video classes, fine bandwidth allocation will be conducted to reallocate the remaining bandwidth to the classes. According to the second principle mentioned above, different priorities are assigned to the video classes: the low-bandwidth class has the highest priority while the high-bandwidth class with the lowest. More specifically, two thresholds T_m and T_l are defined. When the remaining bandwidth is less than T_l, only the low-bandwidth class will be reallocated bandwidth. When the remaining bandwidth is more than T_l but less than T_m, it will be reallocated to the low-bandwidth class and the middle-bandwidth class rather than the high-bandwidth class. The remaining bandwidth will be reallocated to all the classes only when the remaining bandwidth is more than T_m. T_m and T_l can be determined according to the research results that an quality improvement would be visible to the human eyes

only if it is more than 0.5dB in PSNR [24]. In this paper, T_m and T_l are set to be 150Kbps and 50Kbps.

After the fine bandwidth allocation, the output bitrates of the encoders for the three video classes can be adjusted accordingly to match the allocated bandwidth.

(4) Bandwidth allocation when a new client joins.

When a new client C_{new} joins, if there is sufficient bandwidth available for it besides those allocated to the direct-connected clients, C_{new} will be allowed to connect to the video server directly and the allocated bandwidth to other direct-connected clients remain unchanged.,

If the remaining bandwidth is not enough to allocate to C_{new} to make it connect to the video server directly, then:

If C_{new} belongs to the high-bandwidth class, and $n_h = N_{total-h}$ and $N_{total-h} < N_h$ and $N_{total-h} < \dfrac{B_{total}}{B_h}$, then the allocated bandwidth adjustment of all the three classes can be calculated as follows,

$$B_{total} - \alpha(n_h B_h + n_m B_m + n_l B_l) = \alpha B_h \qquad (7)$$

$$\alpha = \frac{B_{total}}{(n_h + 1)B_h + n_m B_m + n_l B_l} \qquad (8)$$

where α is the adjustment factor.

If C_{new} belongs to the mid-bandwidth class, then

$$\alpha = \frac{B_{total}}{n_h B_h + (n_m + 1)B_m + n_l B_l} \qquad (9)$$

And if C_{new} belongs to the low-bandwidth class, then

$$\alpha = \frac{B_{total}}{n_h B_h + n_m B_m + (n_l + 1)B_l} \qquad (10)$$

To avoid severe fluctuation on the bitrate and thus the video quality, a threshold T_{adj} is introduced for α. In this paper, α is set to be 0.9. If $\alpha > T_{adj}$, then the allocated bandwidth of three classes would be adjusted as follows:

$$\begin{aligned} B_h &= B_h \cdot \alpha \\ B_m &= B_m \cdot \alpha \qquad (11) \\ B_l &= B_l \cdot \alpha \end{aligned}$$

If $\alpha \le T_{adj}$ and C_{new} is not the first one in the high-bandwidth class, C_{new} would be specified as a forwarded client which will be discussed in Section IV in detail. If C_{new} is the first one in the high-bandwidth class, then some connected clients would be chosen to disconnect to save enough bandwidth to make C_{new} connect to the video server directly. The rule to choose the client to be disconnected can be determined according to different practical video applications. According to the two principles mentioned above, we assign different priorities to different video clients. The priority of the first client in a class is the highest, and then are the clients in the high-

bandwidth class. The priorities of clients in the low-bandwidth class are the lowest. In such case, we disconnect the video clients with lower priorities.

(5) Bandwidth allocation when a current client disconnects.

If a direct-connected video client disconnects to the video server, then the bandwidth allocated to it would be reallocated to other classes as described in (3).

IV. CLASSIFICATION-BASED MULTI-CLIENT VIDEO TRANSMISSION

In this section, we first proposed a hybrid video transmission scheme by combining Unicast and forwarding, then the proposed Classification-based Multi-client Video Transmission algorithm is presented.

As mentioned above, the output bandwidth of the video server is often very limited. If all the video clients connect to the video server directly and obtain the video streams by Unicast, each video client might not get enough bandwidth to guarantee the reconstructed video quality. What's more, the more the video clients connect, the worse the provided reconstructed video quality will be.

To address this problem, we proposed a hybrid video transmission scheme by combining Unicast and forwarding. More specifically, when there is enough export bandwidth of the server to be allocated or when the adjustment factor $\alpha > T_{adj}$, more video clients will be regarded as direct-connected clients and allowed to connect to the video server directly, and the video server will transmit video streams to these clients by Unicast. But when the remaining bandwidth is not enough to be allocated and the adjustment of allocated bandwidth is bigger than the predefined threshold, i.e., $\alpha \le T_{adj}$, we choose limited video clients to connect to the video server directly as mentioned in Section III, and the others will be regarded as forwarded clients ($\alpha \le T_{adj}$ and C_{new} is not the first), who receive the video streams forwarded from the direct-connected clients in the same classes. The proposed hybrid transmission scheme can efficiently alleviate the pressure on the limited export bandwidth of the video server. It should be noted that the first client of all classes is directly connect to the video server in all cases.

It should be noted that in order to make the video server response faster, the forwarded video clients still send their feedbacks and interactive controls to the video server directly rather than forwarding, because the transmission display of direct-connected clients is smaller than that of forwarded clients. And if a forwarding client disconnects, the forwarded clients would reconnect to the video server automatically. The process of the reconnection is the same as shown in Algorithm 1.

V. PERFORMANCE EVALUATION

In this section, we first evaluate the performance of proposed bandwidth allocation in Subsection A. The performance of the proposed CMVT is then assessed in Subsection B.

The experiments are conducted on a practical video surveillance platform, with multiple clients demanding for the real-time captured and x264 coded video streams through a heterogeneous network connections. The initial bandwidth centers of the high-bandwidth class, mid-bandwidth class and low-bandwidth class are set to be 1300Kbps, 300Kbps and 50Kbps, respectively. The export bandwidth of the video server is set to be 2Mbps, and the video server buffer is set to be 1000K bits, half of the export bandwidth.

A. Performance of Proposed Bandwidth Allocation Algorithm

When multiple video clients with different available bandwidth connect to the video server, in order to evaluate the performance of the bandwidth allocation algorithm, each of the three classes has only one client.

The illustrative network topology is shown in Fig. 3, where VS denotes the video server, A, B and C represents the three video clients, of which the available bandwidth are detected as about 2MKbps, 500Kbps and 128Kbps, respectively.
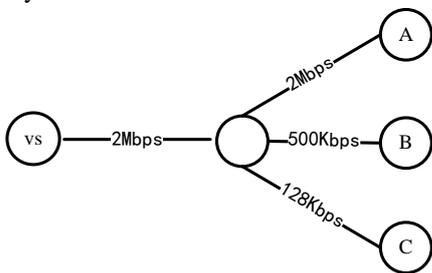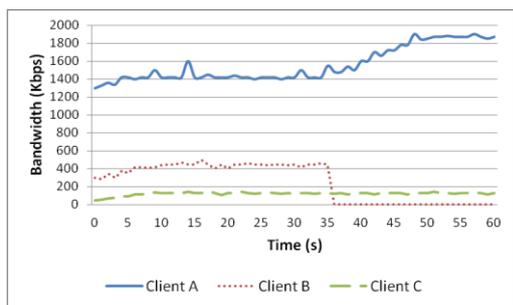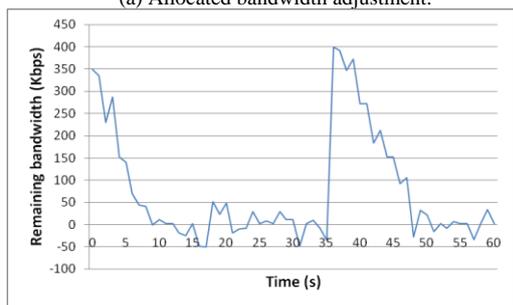

Figure 3.   Clients A, B and C connect to the video server with different available bandwidth.

(a) Allocated bandwidth adjustment.

(b) The remaining bandwidth.

Figure 4.   Bandwidth allocation by our proposed algorithm.

Based on the client classification, Clients A, B and C will be classified into the high-bandwidth, mid-bandwidth and low-bandwidth class, respectively. According to our

proposed bandwidth allocation and video transmission schemes, these three clients will connect to the video server directly. The total bandwidth which has been allocated to the three video clients is 1650Kbps, less than 2Mbps. So the remaining 350Kbps bandwidth will be further allocated to the classes. The bandwidth allocated to the clients is shown in Fig. 4. From Fig. 4 (b), It can be seen that the remaining bandwidth is more than 150Kbps from time t=0 to t=5, so the bandwidth is allocated to all the three clients. After time t=5, the remaining bandwidth is less than 150Kbps and stops to be allocated to Client A, while Clients B and C are still allocated bandwidth. At time t=7, the remaining bandwidth is less than 50Kbps, which stops to be allocated to Client B either. At time t=10, because the bandwidth allocated to Client C has reached 128Kbps, the maximal available bandwidth of Client C in the low-bandwidth class, the remaining bandwidth is allocated to Client B again. At time t=37, Client B disconnects. The bandwidth allocated to Client B is available to be reallocated, so Client A gains additional bandwidth. The remaining bandwidth converges to around zero in a little more than 10 seconds, and the total export bandwidth of 2Mbps has fully and reasonably allocated to Clients A and C, which indicate that the proposed bandwidth allocation algorithm can adaptively adjust the optimal bandwidth allocation under dynamic networks connections and bandwidth conditions.

The performance of the M-LDLF algorithm proposed in [17] is also compared with our proposed scheme. In our experiment, three video encoders in the video server can be seen as three senders, and the export bandwidth of the video server can be seen as the shared resource. In M-LDLF, each client has a video encoder and the R-D models of the current frame are sent to the resource allocation module. However, it does not consider that if the same amount of additional bandwidth is allocated to an encoder, the improvement of reconstructed video quality would be much bigger when the available bandwidth of the corresponding client is lower. In our proposed bandwidth allocation, when the bandwidth is enough to allocate, each video client will gain additional bandwidth to improve its reconstructed video quality. But when the bandwidth is not enough, the bandwidth is allocated to the video clients in low-bandwidth class first to get the biggest improvement of reconstructed video quality by the same amount of remaining bandwidth.
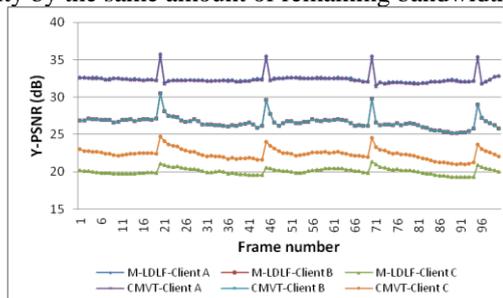

Figure 5.   The Y-PSNR results of reconstructed video when using CMVT and M-LDLF.

Fig. 5 shows the objective quality (Y-PSNR) of reconstructed video provided to the video clients. The

reconstructed video quality provided to Clients A and B are similar to each other using M-LDLF and our proposed scheme. While for Client C, compared with M-LDLF, the average value of Y-PSNR using our proposed scheme increases more than 2dB. This is because for the low-bandwidth class little additional bandwidth can improve the reconstructed video quality greatly, while for the high-bandwidth and mid-bandwidth classes the improvement of reconstructed video quality is negligible by allocating little additional bandwidth. In M-LDLF, three clients compete for the bandwidth all the time regardless of how much bandwidth remains, so Client C cannot get as much bandwidth as in our proposed algorithm for its video transmission, and its reconstructed video quality degrades unavoidably. The experiment results illustrate that our proposed bandwidth allocation algorithm can use the limited bandwidth more efficiently and provide better video quality to clients than M-LDLF.

B. Performance of Our Proposed CMVT Scheme

In this subsection, the experiments are conducted to verify our proposed CMVT scheme. Six Clients A, B, C, D, E and F with different available bandwidth connect to the video server simultaneously demanding for the real-time captured video. The detected available bandwidth of the six clients are 2Mbps, 1.9Mbps, 2Mbps, 500Kbps, 700Kbps, and 128Kbps, respectively.
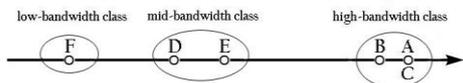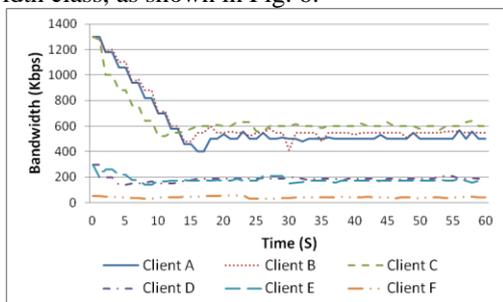

Figure 6.   Clients classification results.

First, based on our proposed client classification scheme, six clients are classified into three classes. More specifically, Clients A, B and C are classified into the high-bandwidth class, D and E are classified into the mid-bandwidth class, and F is classified into the low-bandwidth class, as shown in Fig. 6.

(a) Bandwidth adjustment using M-LDLF.
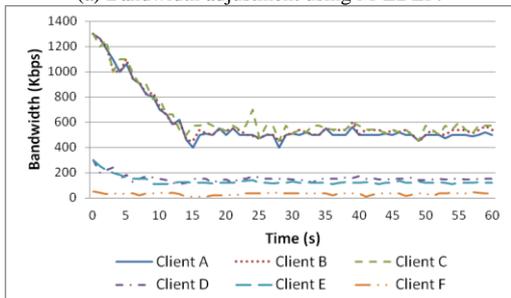
(b) Bandwidth adjustment using Unicast.

Figure 7.   Bandwidth adjustment using different transmission schemes.

Fig. 7 shows the adjustments of allocated bandwidth to six clients when using M-LDLF and Unicast. When using M-LDLF, all the six clients connect to the video server directly, and the video server has to send six video streams to satisfy the requirements of the clients. Although the bandwidth allocator can dynamically distribute the bandwidth to each client, each client can only be allocated much less bandwidth than its available bandwidth because of the limited export bandwidth. The video server has to decrease its output bitrates to enable the video streams to be transmitted under the allocated bandwidth, otherwise a large amount of video packets may be lost and the reconstructed video quality that provided to the clients will be poor. Similarly is the bandwidth adjustment using Unicast.

While in our proposed CMVT scheme, six clients are classified into three classes, only Clients A, D and F connect to the video server directly, and the video server specifies Client A as the forwarding client to forward video to Client B, then Client B forwards video to Client C, and Client D forwards video to Client E. The bandwidth adjustments using our proposed CMVT scheme is similar as shown in Fig. 4 (a), where the bandwidth of Clients A, B and C represents the bandwidth adjustment results of the high-bandwidth class, the mid-bandwidth class and the low-bandwidth class, respectively. The bandwidth adjustment curves of all the three classes increase at the beginning when using our proposed scheme, while using M-LDLF and Unicast the bandwidth adjustment curves decrease at the beginning. This is because the export bandwidth has to be allocated to all the six clients in M-LDLF and Unicast, while in our proposed scheme, the export bandwidth only needs to be allocated to three direct-connected clients, each video client will gain more bandwidth, so the reconstructed video quality provided to all the clients will be better.
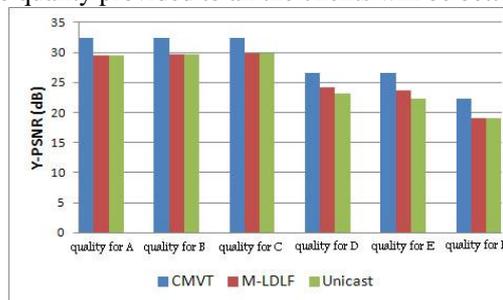
Figure 8.   The reconstructed video quality for six video clients when using CMVT, M-LDLF and Unicast.

The comparison of reconstructed video quality (Y-PSNR) is shown in Fig. 8. As can be seen, the reconstructed video quality using our proposed video transmission scheme is much better than that using the other two algorithms for all the six video clients.

VI.   CONCLUSIONS

In this paper, we have developed a client classification-based bandwidth allocation and video transmission scheme for multi-client video transmission over heterogeneous networks. Multiple video clients are first classified into different classes according to their detected

available bandwidth using discriminant analysis, based on which the limited export bandwidth is allocated to different classes using greedy algorithm. The video streams are then transmitted to the clients using the proposed hybrid video transmission scheme. Extensive experimental results have shown that notable video quality improvements can be achieved for all the video clients with different available bandwidth.

REFERENCES

[1] D. Wu, T. Hou, and Y. Zhang, "Transporting real-time video over the Internet: Challenges and approaches," Proceedings of the IEEE, vol. 88, no. 12, pp. 1855-1877, Dec. 2000.

[2] H. Schulzrinne, S. Casner, and R. Frederick, et al, "RTP: A Transport Protocol for Real-Time Application," RFC 1889, 1996.

[3] Y. Wang, J. Ostermann, and Y. Zhang, "Video Processing and Communications," Prentice Hall, 2002.

[4] H. Kim and S. Park, "A hybrid video-on demand data broadcasting and receiving scheme of harmonic and staggered schemes," IEEE Transactions on Broadcasting, vol.54, no.4, pp. 771-778, Dec. 2008.

[5] Z. Yang, L. Juhn and L. Tseng, "On optimal broadcasting scheme for popular video service," IEEE Transactions on Broadcasting, vol.45, no.3, pp. 318-322, Sep. 1999.

[6] K. Choi, S. Choi, and J. Choi. "Hybrid video transmission for minimizing server bandwidth consumption with zero start-up delay in video-on-demand service," IEEE Communication Letters, vol. 16, no. 1, pp. 6-8, Jan. 2012.

[7] S. Azad and M. Murshed, "An efficient transmission scheme for minimizing user waiting time in video-on-demand systems," IEEE Communications Letters, vol. 11, no. 3, pp. 285–287, Mar. 2007.

[8] W. Shu and M. Wu, "Resource requirements of closed-loop video delivery services," IEEE Multimedia, vol. 11, no. 2, pp. 24-37, Apr.-Jun. 2004.

[9] A. Legout, J. Nonnenmacher and E. W. Biersack, "Bandwidth-allocation policies for unicast and multicast flows," IEEE/ACM Transactions on Networking, vol. 9, no. 4, pp. 464-478, Aug. 2001.

[10] S. Kang and A. Zakhor, "Effective bandwidth based scheduling for streaming media," IEEE transactions on Multimedia, vol. 7, no. 6, pp. 1139-1148, Dec.2005.

[11] K. Nagarajan and G. Zhou, "Self-similar traffic sources: modeling and real-time resource allocation," Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing, pp. 74-77, 2001.

[12] K. Nagarajan and G. Zhou, "A new resource allocation scheme for VBR video traffic source," Conference Record of the 34th Asilomar Conference on Signals, Systems and Computers, vol. 2, pp. 1245-1249, Oct.-Nov. 2000.

[13] Z. Miao and A. Ortega, "Expected run-time distortion based scheduling for delivery of scalable media," The 12th International Packet video Workshop (PVW2002), Pittsburgh, PA, Apr. 2002.

[14] P. Chou and Z. Miao, "Rate-distortion optimized sender-driven streaming over best-effort networks," 2001 IEEE Fourth Workshop on Multimedia Signal Processing, pp. 587-592, 2001.

[15] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," IEEE Transactions on Multimedia, vol. 8, no. 2, pp. 390-404, Apr. 2006.

[16] O. Boyaci, A. Forte, S. Baset and H. Schulzrinne, "Demonstration of vDelay: A tool to measure capture-to-display latency and frame rate," 11th IEEE International Symposium on Multimedia (ISM'09), pp. 444-445, Dec. 2009.

[17] G. Su and M. Wu, "Efficient bandwidth resource allocation for low-delay multiuser video streaming," IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 9, pp. 1124-1137, 2005.

[18] Y. Lin, H. Wu, S. Cheng, W. Wang and C. Wang, "Measuring asymmetric link bandwidths in Internet using a multi-packet delay model," IEEE International Conference on Communications (ICC'03), vol. 3, pp. 1601-1605, May 2003.

[19] A. Nogueira, M. de Oliveira, P. Salvador, R. Valadas, and A. Pacheco, "Classification of Internet users using discriminant analysis and neural networks," Next Generation Internet Networks, pp. 341-348, Apr. 2005.

[20] P. Mahalanobis (1936), "On the generalised distance in statistics," Proceedings of the National Institute of Sciences of India, vol. 2, no. 1, pp. 49-55. 2008.

[21] B. Billah, M.L. King, R.D. Snyder, A.B. Oehler, "Exponential smoothing model selection for forecasting," International Journal of Forecasting, 2006, 22(2):239-247.

[22] S. Everette, Jr. Gardner, "Exponential smoothing: the state of the art-part II," International Journal of Forecasting, vol. 22, no. 4, pp.637-666. 2006.

[23] T. Bergger, Rate Distortion Theory. New Jersey: Prentice Hall, Englewood Cliffs, 1971.

[24] D. Salomon, Data Compression: The Complete Reference (the fourth edition), Springer-Verlag London Limited, ch. 4, 2007.

Bo Li received the B.S. degree in computer science from Chongqing University in 1986, the M.S. degree in computer science from Xi'an Jiaotong University in 1989, and the Ph.D. degree in computer science from Beihang University in 1993.

    Now he is a professor of Computer Science and Engineering at Beihang University, the Director of Beijing Key Laboratory of Digital Media, and has published over 100 conference and journal papers in diversified research fields including digital video and image compression, video analysis and understanding, remote sensing image fusion and embedded digital image processor.

Peng Wang received the B.S. degree from Nanchang Hangkong University, China in 2004 and the M.S. degree from Shandong University, China in 2008, both in computer science and engineering. He is currently working toward the Ph.D. degree in the Department of Computer Science and Engineering at BeiHang University. His research interests are digital video compression and transmission.

Yongfei Zhang received the B.S. degree in Electrical Engineering and the Ph.D. degree in Pattern Recognition and Intelligent Systems from Beihang University, Beijing, China, in 2005 and 2011 respectively.

    Since 2011, He has been an Assistant Professor with the Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, and the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University. He was a visiting research scholar in Video Processing and Networking Lab, University of Missouri, Columbia, U.S, from 2007 to 2009. His current research interests include image/video processing, coding and networking, and video surveillance.

A Simulated Annealing Algorithm for Progressive Mesh Optimization

Junjie Peng, Feixiang Chen[*]
School of Information Science and Technology, Beijing Forestry University, Beijing, China
Email: peng_junjie@126.com, bjcfxiang@163.com

Qi Wang
Graduate School, Beijing Forestry University, Beijing, China
Email: sevenwang@live.cn

Abstract—**This paper presents a novel progressive mesh model based on collapse tree, which can accomplish the disordered expansions and collapses of the simplified mesh and is able to restore the mesh just in a specific area. With arbitrary collapse and expansion, we design an edge collapse algorithm for mesh optimization. Its summary process can be described as follows: generating a series of new simplified meshes through the two operations from the original mesh, meanwhile, controlling the process by simulated annealing algorithm to optimize the mesh constantly, looking forward to a better simplified mesh on global scale.**

Index Terms—**Mesh Simplification; Simulated Annealing Algorithm; Collapse**

I. INTRODUCTION

Mesh model is used to describe the three-dimensional shape of the spatial entity, and is frequently applied to the fields such as geographic information systems, computer graphics, and virtual reality. With the development of science and technology, the acquisition of space 3D data became more and more easily, the spatial mesh data generated by a large number of points is increasingly large. Complex and dense meshes can exhibit entities' physical realism and sense of hierarchy, but the large amount of data brings inconvenience to model storage, transmission, calculation and graphics rendering, especially in the devices with the limit of the hardware conditions [1]. Mesh simplification could not only preserve the original feature of the model, but also perform the entity with the fewer number of meshes, so that the model can meet the users' actual needs and reduce the overhead at the same time.

The most basic problems of mesh simplification are the simplification methods and error metric. Judging from the methods of mesh simplification, the most representative methods are Schreorder's vertex decimation [2], Rossignac's vertex clustering [3], Hoppe's edge collapse [4], and Hamann's triangle

decimation [5]. Vertex decimation is easy to realize, and can guarantee the vertices of the simplified model are the subset of the original model, but it needs to re-mesh the removed polygonal area. Rossignac's scheme is only suitable for the specific algorithm, and cannot guarantee the model topology, while triangle decimation is efficient but cannot be applicable to other meshes, and its implementation is more complicated. Edge collapse is relatively wide used in practice, it can be applied to any meshes with flexible expansibility, and what's more, it is easy to achieve a progressive mesh.

Error metric is to measure the similarity between the simplified mesh and the original mesh, but there is no generally accepted metrics so far. Taking the efficiency and practicality into consideration, general algorithms commonly use partial metric to guide the process of the mesh simplification. Schreorder used the average distance of the point to the plane as a local error metric in his algorithm [2]; Turk took the neighborhood curvature as the measure standard and can maintain the characteristic geometric detail of meshes [6]; Local quadratic error metric method is applied to calculate the approximate error by Garland [7]. Some algorithms, both closely integrated. In some of the algorithms, the two basic problems are closely Contact and blend together, for instance, Eck's simplification model based on wavelet technology [8], and Cohen's envelope mesh simplification model of the global error [9]. Some new scheme introduce concept in other areas to measure the cost [10].

Most existing algorithms [11-17] do the mesh simplification based on the local optimum to current mesh. This could improve the efficiency of the algorithm, however, would result in a cumulative error in the simplifying process, thereby reducing the accuracy of the algorithm. Hoppe gave a progressive mesh optimization method based on energy assessment, it could get a better effect of simplification, also to establish detail mesh model of different levels. However, the algorithm is not efficient, and the progressive mesh model including most other general models comes to a poor effect when establishing models at the level of higher detail in a specific area.

In this paper, we propose a progressive mesh optimization algorithm with less computation and it could preferably achieve the refinement of a particular region based on the collapse tree. General progressive mesh achieve the level of details by recording the relevant information in each step without noticing the relationship between every step, so the progressive mesh is limited to the collapse sequence. In our scheme, not only the collapse information but also the relation information is stored by collapse tree. So collapses and expansions in disorder are available and users can refine a specific area without limits. By introducing the simulated annealing algorithm we optimize the simplified mesh through the two operations, finally resulting in better simplified mesh.

II.   PROBLEM DESCRIPTION

Mesh model contains the information of both data points' space coordinates and network topology: data points are the basic data in mesh model and are also the original data of constructing the network model; network topology information represents the association relationship among vertices, edges and faces, which reflected in graphics is the model's meshes. Thus, the mesh model can defined as:

$$M = (K, V) \tag{1}$$

where V represents the set of vertices and $V = \{v_1, v_2 \dots v_n\}$, K is topology information of mesh model.

Mesh simplification can be divided into two situations, one is simplifying in a certain residual threshold, and the other is under a certain compression ratio. Due to error metrics in vary algorithms are different and there is no unified standard, the application of first situation is restricted to the specific algorithms. So in this paper, we choose the latter to describe our mesh simplification.

Mesh simplification is to reduce meshes (or vertices) to the greatest extent while preserving the original mesh shape features as many as possible [18]. When the number of vertices is fixed, determining which simplified mesh is better should measure the matching degree between simplified mesh and original mesh. Therefore, we need to design a simplification method and establish a matched method of error metrics as well. By simplifying the mesh according to the error metrics, more satisfied meshes are obtained.

III.   PROBLEM SOLUTION

A. Method of Mesh Simplification

There are many ways to simplify meshes and different methods have different advantages and disadvantages as

we have discussed. Taking the expandability and simplicity into consideration, we finally adopt the edge collapse as the basic operation for our simplification process. Any meshes such as quad meshes [19] can be subdivided to triangle meshes, to describe the algorithm conveniently and uniformly, we take triangle mesh as the example, and introduce the basic concepts.

(1) Collapse point

Collapse point refers to the new vertex generated in the process of mesh simplification. During multiple collapse processes, collapse point can be collapsed again. Hence, the collapse point can be divided into two kinds. One is visible collapse point, that hasn't been collapsed again after being created, and it is one of vertices in the current mesh. In the view of users it is visible. The other kind is intermediate collapse point, which has been collapsed more than twice and is invisible. Visible collapse point and invisible collapse point is relative to a specific mesh. As shown in Fig. 1 collapse point v_{12} is visible in M_2 (v_{12} is the vertex of M_2), while it is invisible in M_3. Therefore, for a certain simplified mesh, the vertices can be divided to original data vertex and visible collapse point, let primitive mesh model as $M_0 = (K_0, V_0)$, then any simplified mesh can be represented as:

$$M = (K, V_s, V_v) \tag{2}$$

where $V_s \in V_0, V_s \bigcup V_v = V$ and V_v is the set of visible collapse vertices in M.

(2) Relevant vertex

In the process of collapses, it is necessary to record the relevant vertex of collapse edge in order to reversed recovery which is the base of progressive mesh. Relevant vertex is another vertex that constitutes triangle besides the two of collapse edge. Generally, collapse edge that is not the border corresponds to two relevant vertices, while the border corresponds to only one relevant vertex. As shown in Fig. 2, the relevant vertices corresponding to collapse point v12 are v_6 and v_7. Hence, let collapse vertices be v, the vertices been collapsed are v_1 and v_2, the relevant vertices r_1 and r_2, thus one collapse can be noted as:

$$v(v_1, v_2, r_1, r_2) \tag{3}$$

(3) Collapse tree

In our scheme, the collapse tree is the basic unit to record collapse operations. Its data structure is binary tree owed to its name. For any visible collapse point $v \in V_v$ in simplified mesh, there is a corresponding collapse tree, such as v_{12} in M_2, v_{123} in M_3 shown in Fig. 2.
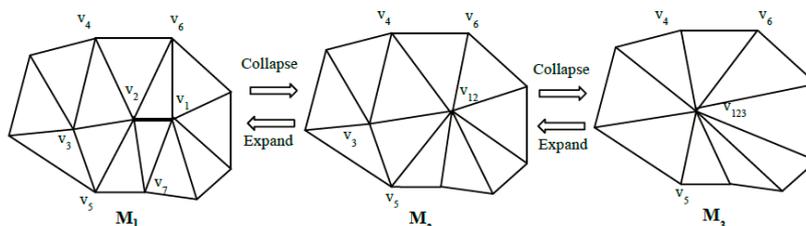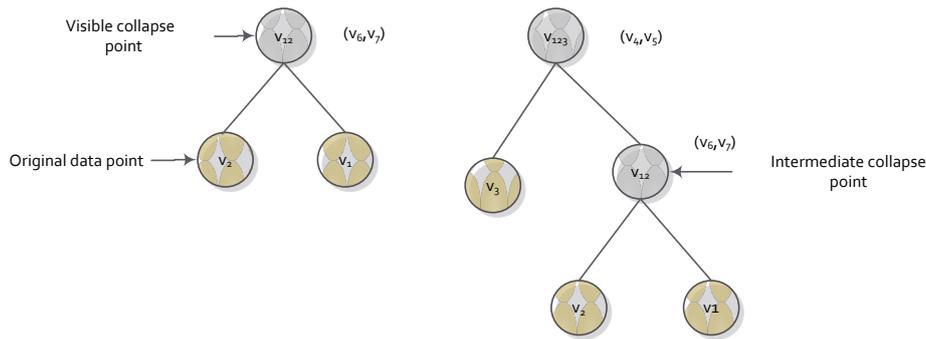

Figure 1.   Collapse and Expansion

Figure 2.   Collapse Tree

The pairs beside the collapse vertices are relevant vertices in Fig. 2, particularly for the collapse vertex that has only one relevant vertex the vacant position puts 0 instead. The two child of any non-leaf node corresponds to the two vertices of collapse edge, and every collapse operation equals two collapse trees correspond to the two vertices of collapse edge composes to a new binary tree, and the tree root is the new vertex after collapse. So every visible collapse vertex has a collapse tree rooted with it. We use $T(v)$ to represent the collapse tree rooted with v. Thus, the results of this progressive simplification are the simplified mesh and the collapse trees of the mesh's visible collapse points, which store the collapse information of the simplification process.

For a collapse tree, we can get its leaf nodes which are the vertices of the original mesh noted as S by tree traversal. Then it is easy to obtain the relevant area Star(S) of the original mesh from the points set S. while with the collapse tree root v, we can get the relevant area Star(R) of the current mesh, then calculate the local error between the Star (S) and Star(R) to regional mesh. The error is called collapse error $ET(v)$ and we use it to represent the collapse error of the collapse tree rooted with v, in particular $ET(v_0 = 0)$ if $v_0 \in V_0$.

B. Expansion of Simplified Mesh

Progressive mesh [20] can be recovered though the reversed operations based on the recorded collapse information. In this paper, the reversed operation is called expansion. For a general collapse progressive mesh model, the model can be expanded successfully only through expansions strictly in accordance with the order of collapse. However, with the extra information stored in collapse tree, our simplified mesh can be expanded in disorder

The main factors to block the expansions in disorder are the changes of relevant vertices. The relevant vertices recorded in the process of collapse will change during the process of collapse and expansion, expansions in order can ensure that the relevant vertices can be restored in sequence, so that it can guarantee the accuracy of the relevant vertices in every expansion operation. But if being expanded disorderly this can't be guaranteed.

Therefore, in order to realize the disordered expansion, we need to solve the two problems: (1) How to update the information of the collapse points when its relevant vertices are collapsed in the later process; (2) How to deal with the change when the relevant vertices are expanded.

To the first problem, according to the collapse tree we can track back to the visible point in current mesh from the original relevant vertex and replace it. As Fig. 2, the relevant vertex of v_{123} is v_4, and we can track back to v_{46} through the collapse tree, in this way the mesh can be expanded successfully.

The solution to solve the second problem is to update the information about relevant vertices of the collapse vertex. For example, when collapse point A (A_1, A_2, B, C) is expanded, we should update the collapse point D (D_1, D_2, A, E) of which A is one of the relevant vertex. The specific updating method should be discussed in the two situations:

(1) In the two vertices generated by the expansion of collapse point A, there is only one vertex that is linked with D, let the vertex be A_1, then the relevant vertex A should be updated as A_1.

(2) When the collapse point D is linked with the two vertices generated by expansion of A, the situation will be more complicated. The brief derivation is given as follows. According to the assumed condition, D is linked with both A_1 and A_2, then there must be one vertex among A's relevant vertices B and C being D's expansion vertices (D_1 or D_2) or the ancestor node of D's expansion vertices. Suppose Ps (B) = D_1 (Ps (P) stands for any node on the vertex P's root path in the collapse tree). If D is expanded at present there will be an ambiguity as shown in figure 3, but as Ps (B) = D_1 and B is the relevant vertex of A, D_1 is linked with A_1, A_2, situation b can be eliminated. That is to say, D's relevant vertex A should be updated to A_2, which is linked to D_2.
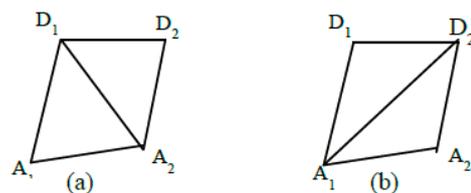

Figure 3.   Two ways of expansion

Taking the mesh M_3 as an example in Fig. 1, the edge $v_4 v_6$ is folded to v_{46} (v_4, v_6, v_{123}, 0), and then we firstly expand v_{123} and then v_{46} instead of the collapse sequences, as follows in Fig. 4.
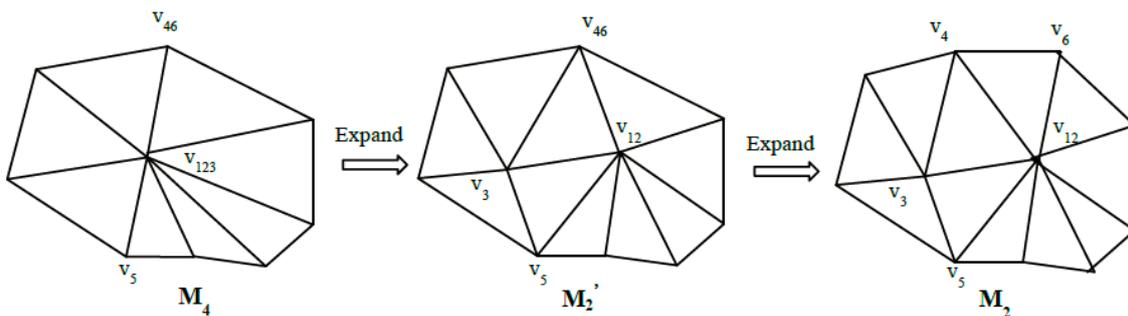
Figure 4.   Expansion in disorder
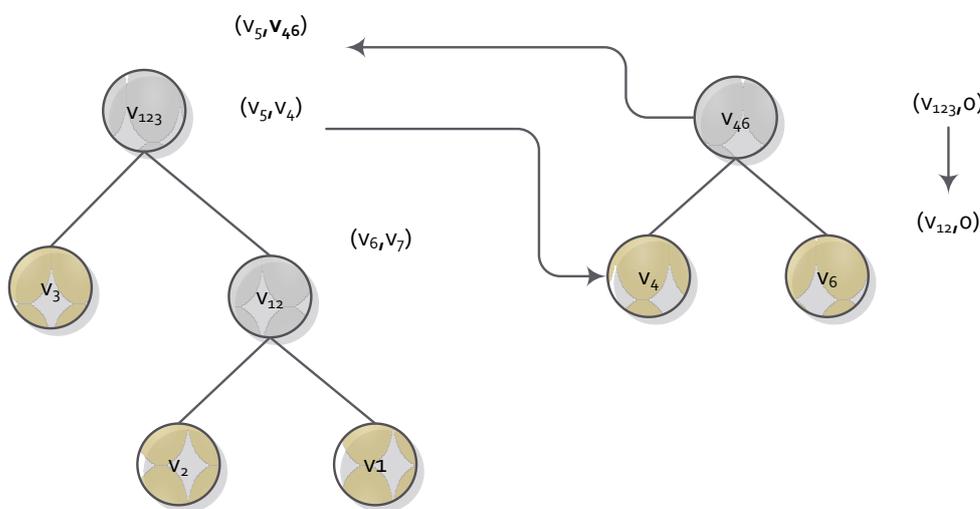
Figure 5.   Trace collapse information

When expanding v_{123} (v_{12}, v_{13}, v_4, v_5), there is a problem that v_4 is no longer in mesh (v_4 is invisible). But with the collapse tree, we can track back to root node corresponding to v_4, then expand it, and get mesh M_2', meanwhile update v_{46} (v_4, v_6, v_{123}, 0), which takes v_{123} as relevant vertices. It's easy to   reckon that v_{12} is linked to v_4 and v_3 according to the rules above, so we choose v_{12}, connected to v_6, to replace v_{123}, thus M_2' can be restored to M_2 by expanding v_{46}, and then M_2 can be restored to the original mesh by spreading v_{12}.

C. Progressive Mesh Model Based on Collapse Tree

Based on the theoretic foundation, we can build a progressive mesh model based on collapse tree. The model takes the collapse points as the base to store the collapse information and the hierarchical relation between the collapse points, which makes it possible to implement the disordered expansions of any visible point in simplified mesh, and the algorithm in the following section is based on this.

Generally, the algorithms just use collapse steps to store the information of progression. Those mesh model can be expanded successfully by inversing the collapse sequence. However they are inappropriate to the expansions just in one certain area. For example, in the progressive transmission, detail presentation and virtual

reality in some limited situation [21], we need the detailed information of certain area, and then we cannot but traverse most collapse information in other areas as well which is actually unnecessary. Our mesh model, by contrast, can expand the mesh by choosing collapse points in certain areas and avoid redundant transmission and operations.

D. Measurement of Mesh Error

Error measure is used to quantify the difference between the input model and the output model. We need the difference to tell which simplified mesh is better. So how to measure the match degree between the simplified mesh and the original is one of the key problems that should be solved and what's more the error metric is the basis of simplification, and is used to guide the simplification process.

There are several schemes to measure the error. One simple way can be defined as follows, as the distance between corresponding sections of the meshes.

$$E(M_1, M_2) = \max_{v \in M_1} d_v(M_2) \qquad (4)$$

where $d_v(M)$ is the distance from vertex v to mesh M, $d_v(M) = \min_{w \in M} ||v - w||$ ($|| \cdot ||$ means   the   Euclidean distance of the two vectors).

Sometimes we need a mean result, with the distance define above, it's easy to get the mean distance.

$$E(M_1, M_2) = \frac{1}{|M_1|} \int_{M_1} d_v(M_2) ds \qquad (5)$$

Note that this definition of distance is not symmetric, that is to say $E(M_1, M_2)$ may not equals $E(M_2, M_1)$. So in this paper, we adopt a two-sided distance (Hausdorff distance) to measure the difference between the two mesh models [22]. Let $E(M_1, M_2)$ be the distance between mesh M_1 and M_2, then

$$E(M_1, M_2) = \max(\max_{v \in M_1} d_v(M_2), \max_{v \in M_2} d_v(M_1)) \quad (6)$$

E (M_1, M_2) is used to measure the maximum deviation between the two models in our scheme.

IV. ALGORITHM DESCRIPTIONS

Simulated Annealing Algorithm (SAA) is a calculation model simulating the annealing process of objects. It is first proposed in 1983 by S. Kirkpatrick, et al. [23], and it is a widely used random search algorithm for solving global optimization problems [24, 25]. Simulated annealing algorithm is different from the traditional random search strategy, it not only introduces the appropriate random factors, but also borrows the natural mechanism of the annealing process in the physical system [26], and consequently there is a greater probability of global optimal solution. Therefore, we apply SAA to the mesh simplification problem to get a global optimal mesh or better simplified mesh from a global perspective. Combining the principles of the general simulated annealing algorithm with mesh simplification, the proposed algorithm consists of the following steps:

Step 1: (Data Preparation) Input the original mesh M_0. Calculate the number of collapses according to the compression rate.

Step 2: According to the number of collapses, collapse the original mesh M_0 randomly to generate the initial mesh as the state of the initial solution M, and in the process record each collapse step in accordance with the proposed method. Set the parameters, including the initial temperature T (sufficiently large), and the iterations L for every T.

Step 3: In the mesh of current solution, take a visible collapse point randomly to do the expand operation, then select one edge to fold, so that a new solution M' could be produced.

Step 4: Calculate the energy difference ∆E before and after step 3.

Step 5: If the energy is reduced (∆E<0), accept the new solution as current solution, otherwise, accept M' as the current solution with the probability of $e^{-\Delta E(M_1, M_2)/T}$.

Step 6: If the iterations does not exceed the setting and the accuracy does not meet the requirements, turn to step 4.

Step 7: Finish. Output the current solution as the result.

A. Generation of Iinitial Mesh

Our simplification is under a certain compression rate, so the number of faces in the ultimate mesh could be calculated by the compression rate and the original mesh. Then it's easy to get the initial mesh for the algorithm containing specific number of faces. The initial mesh is generally not optimal, and we used it as the initial solution of the simulated annealing algorithm. There are two solutions to generate initial mesh, one is to do the collapse operation on the edge randomly; the other to get the initial mesh is by the existing schemes of mesh simplification, then use SSA to optimize the initial mesh, something like the scheme in [27, 28]. According to the compression proportion, the initial mesh is generated after certain collapses from the original mesh, and the collapse information required by the progressive mesh model is recorded as well.
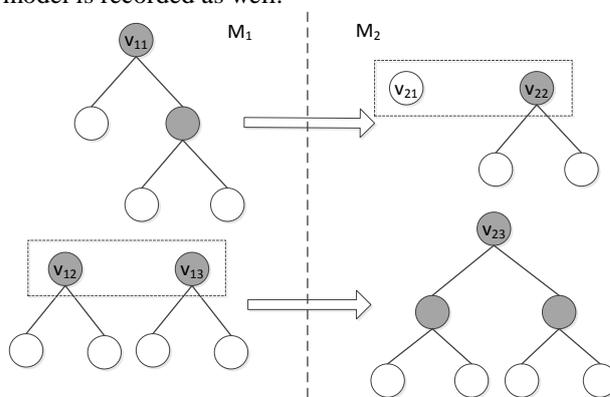
Figure 6. The changes after collapse and expansion

B. Transition of State

In order to ensure the certain number of triangle faces in the mesh (i.e. the compression rate remains unchanged), we do expand and collapse operation respectively for one time to the current mesh as the state transition in SSA. The concrete steps are: select a visible point from the collapse points in current mesh to do expand operation, and then select the one edge of the mesh to collapse. For the new solution, we need to determine whether it is acceptable or not based on simulated annealing algorithm. In order to avoid large number of repetition calculations, the basic method is compared the difference between the two meshes. The normal algorithms are just to compare the relative error between the current mesh and the former mesh, so the difference error is shown as:

$$\Delta E(M_1, M_2) = E(M_2, M_1) \qquad (7)$$

However, the method would lead the error to be accumulated. So we firstly compare the current mesh and the former to the original mesh, and then use their difference value.

$$\Delta E(M_1, M_2) = E(M_2, M_0) - E(M_1, M_0) \qquad (8)$$

The amount of the theoretical error calculation is very large, so in the actual application, representative sample points in both mesh surfaces are usually used to do
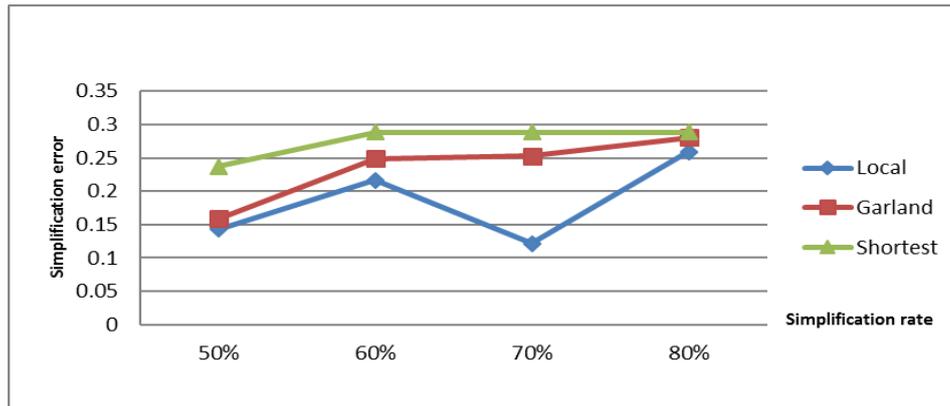
Figure 7.   The Compare results of the model cow
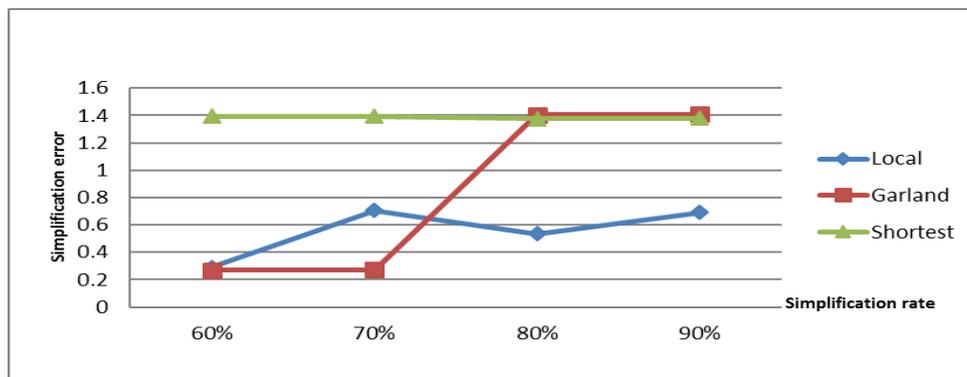
Figure 8.   The Compare results of the model Porsche
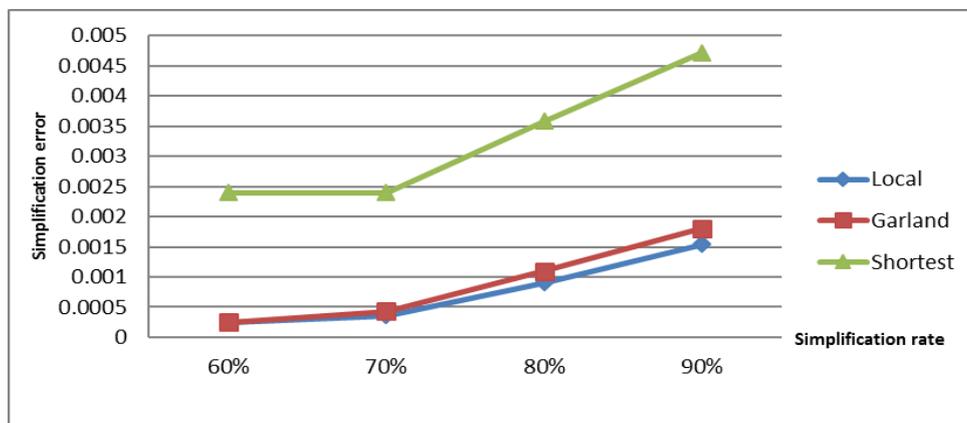
Figure 9.   The Compare results of the model bunny

approximate calculation (these sampling points including at least the vertex of each mesh). The same approximate method is adopted in this paper. The process of state transition will involve six collapse trees (including the original data points). The Fig. 6 shows the trees of both former state and new state respectively.

TABLE I.   BASIC DATA OF THE TEST MODEL

Model	Vertex	Face	Area	Diagonal
Cow	2903	5804	21.1821	5.603207
Porsche	5247	10474	276.7388	15.814117
Bunny	35947	69451	0.0571	0.250246

V.   EXPERIMENTAL RESULTS AND ANALYSIS

In order to test the effect of our simplification algorithm, we do some comparison test between our algorithm and some other classical algorithms (including the Garland method and the Shortest Distance method). The error is calculated using the Metro. Metro is a tool designed to evaluate the difference between two triangular meshes, and it adopts an approximated approach based on surface sampling and point-to-surface distance computation. This experiment is carried out without introducing a new point, and therefore the

collapse point can only choose one from the two vertices of the collapse edge with smaller error.

In the experiments we have carried out the tests with three models respectively, the basic data of the three models are shown in Tab. 1. For each model, the comparisons of the three simplification algorithms are carried out at the different compression ratios. The results are shown in Fig. 7, 8, 9.

From the results, we can conclude that our algorithm has lower error relative to the other algorithms at higher simplification rate. In addition, the algorithm can be used as a method of optimization to other algorithms. For a more detailed explanation, we use other algorithms to simplify the model at the fixed compression ratio, take the result as the initial solution, and then optimize it with our algorithm, thereby to obtain more preferably simplified mesh.

VI. CONCLUSION

The progressive mesh model based on collapse tree can collapse and expand the mesh disorderly. Depending on this characteristic, it is able to convert the mesh optimization problem to the combination optimization problem. In this paper, we adopt the simulated annealing algorithm to search the more optimal simplified mesh on global scale. During the optimization process, by tracking the relevant information of the original mesh according to collapse tree and calculating the error with the original information, we can avoid progressive deviation caused by calculating the current mesh merely. Moreover, taking the visible collapse points as the expansion unit can meet the requirement of refining certain areas in simplified model, which can be used to improve the mesh's progressive transmission and multiple-level detail model.

REFERENCES

[1] Y. P. Zhang, H. Xiong, et al. "A survey of simplification and multiresolution techniques for massive meshes", *Journal of Computer-Aided Design & Computer Graphics*, 2010.

[2] W. Schroeder, J. Zarge, W. Lorensen, "Decimation of triangle meshes", In: *ACM SIGGRAPH Computer Graphics*, vol. 6, no. 2, pp. 65-70, 1992.

[3] J. Rossignac, P. Borrel, "Multi-resolution 3D approximations for rendering complex scenes", In: *Modeling in Computer Graphics, Eds.Springer-Verlag*, pp. 455-465, 1993.

[4] H. Hoppe, T. DeRose, T. Duchamp, et. al. "Mesh optimization", In: *Proceedings of the 20th annual conference on Computer graphics and interactive techniques. ACM*, vol. 27, pp. 19-26, 1992.

[5] B. Hamann, "A data reduction scheme for triangulated Sur-faces", *Computer aided geometric design*, vol. 11, no. 2, pp. 197-214, 1994.

[6] G. Turk, "Re-tiling polygonal surfaces". In: *ACM SIGGRAPH Computer Graphics*, vol. 26, no. 2, pp. 55–64, 1992.

[7] M. Garland, P. S. Heckbert, "Surface simplification using quadric error metrics", In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co*, pp. 209-216., 1997.

[8] M. Eck, T. DeRose, T. Duchamp, et al. "Multiresolution analysis of arbitrary meshes", In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. ACM*, pp. 173-182, 1995.

[9] J. Cohen, A. Varshney, D. Manocha, et al. "Simplification envelopes", In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. ACM,* pp. 119-128, 1996.

[10] L. P. Xing, K. C. Hui, "Entropy-based mesh simplification", *Computer-Aided Design and Applications*, vol. 7, no. 6, pp. 911-918, 2010.

[11] Z. G. Pan, K. Zhou, J. Y. Shi, "A new mesh simplification algorithm based on triangle collapses", *Journal of computer science and technology*, vol. 16, pp. 28, 2001.

[12] X. L. Liu, Z. Y. Gao, D. Peng, et al. "Collapse simplification based on sharp degree", *Journal of software* vol. 16, pp. 669-675, 2005.

[13] J. Wang, M. Y. He, "An algorithm for mesh simplification based on the importance of vertex and its application", *Application research of computes* vol. 21, pp. 41, 2004.

[14] M. Garland and P. S. Heckbert, "Simplifying surfaces with color and texture using quadric error metrics," In: *Visualization '98*, 1998.

[15] D. Luebke, C. Erikson, "View-dependent simplification of arbitrary polygonal environments" In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques.* ACM Press/Addison-Wesley Publishing Co., pp. 199-208, 1997.

[16] F. Wan, F. He, et al. "A simplification algorithm based on appearance maintenance", *Journal of Multimedia*, vol. 5, no. 6, pp. 629-638, 2010.

[17] S. Zhang, J. Zhao, et al. "Feature aware multiresolution Animation Models Generation", *Journal of Multimedia*, vol. 5, no. 6, pp. 622-628, 2010.

[18] Z. G. Pan, X. H. Ma, J. Y. Shi, "Overview of multiple level of detail creation", *Journal of image and graphics* vol. 3, pp. 754-759, 1998.

[19] M. Tarini, N. Pietroni, P. Cignoni, et al. "Practical quad mesh simplification", *Computer Graphics Forum*, vol. 29, no. 2, pp. 407-418, 2010.

[20] H. Hoppe, "Progressive meshes", In: *Proceedings of the 23rd annual conference on computer graphics and interactive techniques. ACM, 1996.*

[21] N. Aspert, D. Santa-Cruz, T. Ebrahimi, "Mesh: Measuring errors between surfaces using the hausdorff distance", In: *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference*. IEEE, pp. 705-708, 2002.

[22] S. Kirkpatrick, D. G. Jr., M. P. Vecchi, "Optimization by simulated annealing", In: *Science,* vol. 220, no. 4598, pp. 671-680, 1983.

[23] A. Dekkers, E. Aarts, "Global optimization and simulated annealing" In: *Mathematical programming*, vol. 50, pp. 367-393, 1991.

[24] L. Ingber, "Adaptive simulated annealing (ASA): Lessons learned", Control *and cybernetics,* 1996.

[25] T. Yin, Y. Han, Y. Chen, et al. "WebVR——Web virtual reality engine based on P2P network", *Journal of Networks*, vol. 6, no. 7, pp. 990-998, 2011.

[26] R. L. Yang, J. F. Gu, "An Efficient Simulated Annealing Algorithm for Global Optimization", *Systems engineering-theory & practice*, vol. 17, no. 5, pp. 30–36, 1997.

[27] V. Vidal, C. Wolf, F. Dupont, "Combinatorial mesh optimization", *The Visual Computer*, vol. 28, no. 5, pp. 511-525, 2012.

[28] L. Chen, M. Holst, "Efficient mesh optimization schemes based on optimal Delaunay triangulations", *Computer*

Methods in Applied Mechanics and Engineering, vol. 200, no. 9, pp. 967-984, 2011.

Junjie Peng was born in the city of Enshi, Hubei Province, China, in 1987. He graduated and received his bachelor degree in management from the information management and information system major in Beijing Forestry University in 2011. Currently he is a second-year postgraduate of the same school and his main research field is topology generation in GIS.

Feixiang Chen was born in Hubei Province of China in 1977. He received the M.S. degree in computer applied technology from China University of Geosciences in 2003, and received the Ph.D degree in cartography and geographic information system from Chinese Academy of Sciences in 2006. Now he is an associate professor in Beijing Forestry University, and his main research field is mobile GIS.

Qi Wang was born in the city of Hangzhou, Zhejiang Province, China, in 1989. He graduated and received his bachelor degree in management from the information management and information system major in Beijing Forestry University in 2012. Currently he is a first-year postgraduate of the same school and his main research field is Mobile GIS.

Crowd Density Estimation Based on Texture Feature Extraction

Bobo Wang, Hong Bao*, Shan Yang, and Haitao Lou

Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing Chaoyang district, 100101, China

Email: hbcxwtb@gmail.com, baohong@buu.edu.cn, 772817987@qq.com

Abstract—As we know, feature extraction has an important role in crowd density estimation. In our paper, we introduce a new texture feature called Tamura, which is usually used in image retrieval algorithms. On the other hand, the time consuming is another issue that must be considered, especially for the real-time application of the crowd density estimation. In most methods, multiple features with high dimension such as the gray level co-occurrence matrix (GLCM) are used to construct the input feature vector, which will decrease the performance of the whole method. In order to solve the problem, we use Principal Component Analysis (PCA) method, which can obtain the mainly information of the feature using less dimension features. In the end, we use the Support Vector Machine (SVM) for estimating the crowd density. Experiments demonstrate that our method can generate high accuracy at low computational cost compared with other existing methods.

Index Terms—GLCM, Tamura texture features, dimensionality reduction, SVM

I. INTRODUCTION

With the rapid economic development and increasing people's social activities, the flow density of large malls, supermarkets and places such as subway station is growing more and more serious and brings security risks by the crowd congestion. According to news agency reports: On February 27, 2013, LaoHeKou of Hubei province occurred a stampede caused by crowd of local elementary school, 11 students were injured, 4 students died. Therefore, if we can analyze the events in advance and predict the likelihood of the risk, then timely and effective measures can be took to ease the crow. It can effectively avoid the social security problem that caused by the high crowd density. According to the different feature extraction methods, crowd density estimation techniques can be divided into three categories: 1) estimation based on pixels statistical categories; 2) estimation based on texture analysis technique;3) estimation based on individual characteristics. For the first category, Davies, Chow et al proposed a approach of using foreground-pixels, which is based on the relationship between number of people and crowd density

using a linear methods for estimating the density. The Hong Kong Chow et al proposed a neural network method by using a mix of global learning algorithm for the crowd density. The method is mainly based on the extraction of the three characteristics of the crowd: population object edge length, a proportion of object pixels in the image and the background of the object pixels in the image. Paragios [1] proposed a method based on Markov Random Field (MRF) real-time crowd density estimation. The method is divided into two steps: The first step is to distinguish the image's foreground and background by Markov Random Field; The second step is to get the image by changing detection combined with a geometric module to perspective correction, and then estimate the population density of the monitored area. Chow et al. joined the body template matching module and match for human body for the people who are clearer individual characteristics and near camera. The performance of the system has been obviously improved and is applied in the subway system in Hong Kong. In the method of crowd density estimation based on texture analysis technology: it was proposed by Brazilian Marana as early as in 1998. Marana think that images of crowds with different densities tend to present distinct texture patterns. High density crowded areas are often made up of fine patterns (which correspond to high frequencies in the frequency domain), while images of low density crowded areas are mostly made up of coarse patterns (which corresponds to low frequencies in the frequency domain), especially when their backgrounds are also made up of coarse patterns. H. Rahmalan [2] take the crowd image as texture image, and extract gray level co-occurrence matrix and the min Khodorkovsky fractal dimension, scale invariant orthogonal Chebyshev torque characteristics. Finally, crowd density is classified by using self-organizing mapping learning and classification algorithms. Wu [3] segments image into different sizes by adopting the method of projection transformation in accordance with the different distances from the camera. Then extract texture feature vectors from each input image grid and use support vector machine (SVM) method to get the population density in each grid. Finally, the support vector machine (SVM) method id used again to detect abnormal density distribution based on the estimated density vector. CHAN [4] established the regression equation by comprehensive utilization the

---

groups of pixels and texture feature information, and get the function relation the number of scenarios and the characteristics of the crowd. The Qing WEN et al [5] proposed the mathematical method by the texture characteristics of the crowd and the number of the scene to statistics the number. The method of using the 2-d Gabor filter to extract the crowd global texture characteristics can effectively avoid the overlap between the crowd and perspective distortion. In paper [6], the authors put forward a new kind of approach which through the accumulation of Mosaic image difference feature to represent these complex random motion models for accurate detection. Then, through the perspective distortion correction model, the observed field achieved considerable population density measurements obtained on the basis of the foreground image. Chow methods cannot solve the problem of high density crowd to distinguish, but crowd density estimation method based on texture analysis technology can solve it. This method is currently under further development. In crowd density estimation based on individual characteristics, Shcngfim Lin etc. put forward a kind of through the crossing transform to extract the head contour feature to locate each one, to determine the number of a crowded environment. D.B. Yang [7] obtained images of the same scene by multiple sensors from different angles computes scenario 2d projection and used background difference method to detect prospect. Then, prospects are detected from various angles for fusion. Finally, geometric algorithms are used to calculate the projection area of the body contour, and count. Antonio etc. [8] proposed a new feature which is used for people counting and statistical method, which made use of feature points to count the number. They assumed that the number of feature points and the crowd density and the quantity have some corresponding relation. The method achieved ideal effect, and some gesture recognition is studied by using characteristic points at the same time. Conte [9] classified the feature point of prospects for detecting movement in the scene. Then the feature points in each category, the number of parameters such as the distribution of feature points were trained in SVM for ultimately determining the number of pedestrians.

Currently, many researchers who are devoting the crowd behavior analysis usually use the PETS data sets [14]. The PETS (Performance Evaluation of Tracking Systems) meeting is AVSS BBS annual meeting on crowd behavior analysis. In recent years, the meeting has been held in the United States, Britain, Australia, Italy and many other countries. The meeting is mainly on crowd behavior analysis, including population statistics, crowd density estimation, individual staff and tracking individuals in the group, special groups and special events detection, etc. The PETS data set includes the background and normal flow for the training data, and the density estimation data, the data used for pedestrian tracking and behavior analysis and event detection.

This paper is a new kind of texture feature of crowd density estimation. Tamura feature is a kind of common texture feature in image retrieval. Through the feature extraction and PCA dimension reduction [10], the obtained feature vector has high correlation with crowd density, which can improve the accuracy of crowd density estimation.

The rest of this paper is organized as follows. In Section II, the details of our proposed method are elaborated. Section III demonstrates the experiment results. Finally, some concluding remarks are presented in Section IV.

II.   PROPOSED METHOD

The flowchart of our proposed method is illustrated in Figure. 1. We can see that, the images are firstly preprocessed, which mainly includes: get gray image, deal with the noise, and extract the interested area. Secondly, the 13 image texture features are extracted in the area of interest by gray level co-occurrence matrix. Then dimension reduction is implemented by principal component analysis (PCA) to get the high correlation of 4-D feature vector. Four Tamura image texture features are extracted by Tamura. Through feature fusion, an 8-dimensional feature vector is got. Finally, classification and prediction are obtained by using the SVM [11] to get the estimation of crowd density.
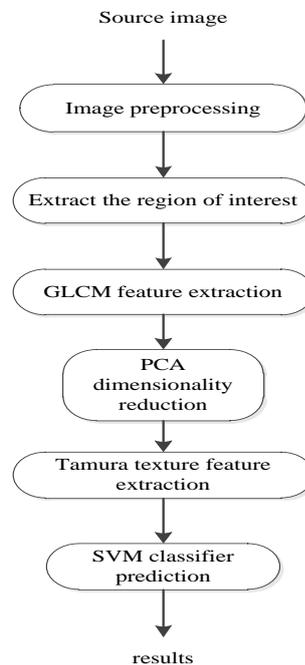
Figure 1.   Flowchart

A. The Image Preprocessing and Interest Region Extraction

Usually, there are laws in the pedestrian route of movement in the scene, namely, they walk along a certain route. For this experiment, the road that the pedestrians are walking on is the area of interest [12]. After picking up the area of interest, we can remove the influence caused by the background in the image, which can improve the estimated accuracy and efficiency. Figure 2 are the area of interest extracted by graying image and Wiener filter denoising for each set of source images.

(a)

(b)

(c)

(d)
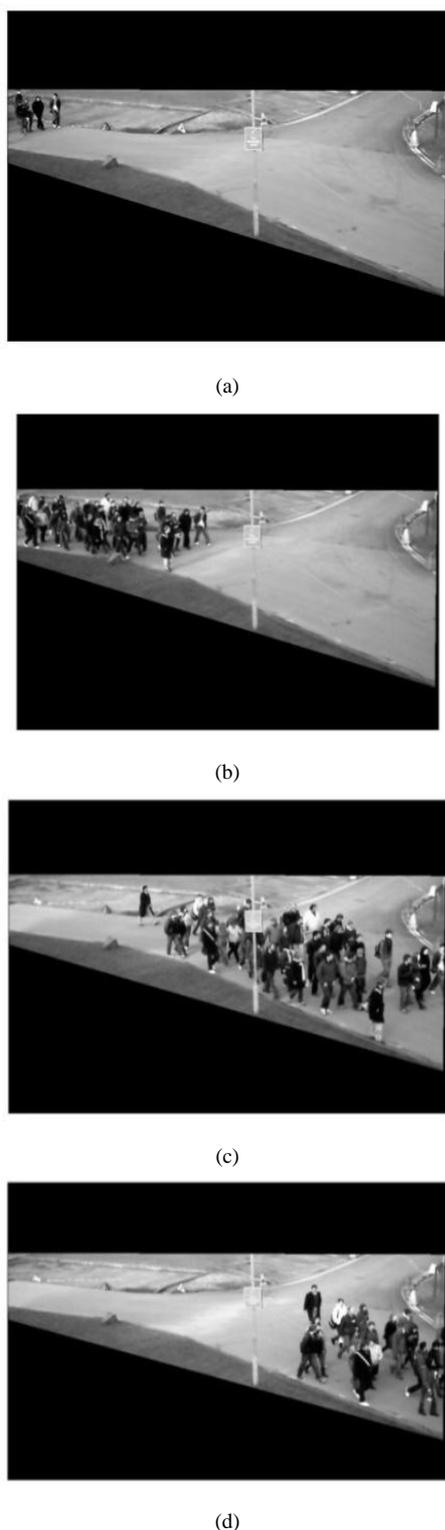
Figure 2. (a), (b), (c), (d) are the area of interest extracted by graying image and Wiener filter denoising for each set of source images.

B. GLCM Texture Feature Extraction and Dimension Reduction

In texture feature extraction [13], Marana used GLCM texture analysis methods. The method calculates GLCM of the crowd image and extracts the texture feature. Then the feature vectors into neural network classification.

Because the different density of the population images corresponds to a different texture pattern, in the high density of population, the texture is a fine and coarse mode in the low-density populations' image.

Texture is alternately repeated by the gray scale distribution on space position change. So there must be some gray relation between two pixels separated by a distance in the image space. This is called the image gray-scale spatial correlation properties [14]. Through the study of spatial gray level correlation to describe texture, this is the thinking foundation of gray level co-occurrence matrix. Gray level co-occurrence matrix is the probability of i point to the gray j point and the distance is $d = (Dx, Dy)$ .Gray level co-occurrence matrix express with $P_d(i,j)(i, j = 0,1,2,...L-1)$, which L is the gray scale of the image, i, j, respectively pixel gray level and d is the space position relation between two pixels. Different d determines the distance and direction between two pixels. Θ is the generating direction of gray level co-occurrence matrix. Usually the four directions are $0^0,45^0,90^0$ and 135^0. As shown in figure 3.
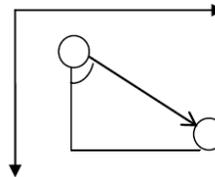
Figure 3. Pixel gray level co-occurrence matrix

When d is selected, it generates gray level co-occurrence matrix. One element of the co-occurrence matrix represents the number of occurrences of combinations of a gray scale. In general, the gray scale of a image is 25.When calculating the texture feature which are from the the gray level co-occurrence matrix, it request the gray scale image is far less than 256. That is mainly because of matrix dimension is larger and the size of the window is smaller. If the texture is very good to describe, window size is larger, and it can greatly increase the amount of calculation. Furthermore, each type of boundary areas of the error rate is bigger when the window size is big. So we need do image histogram specification before calculation of gray level co-occurrence matrix to reduce the image gray scale. General the gray scale of a image is 8 or 16 after image histogram specification.

In this paper, we extract texture features from GLCM. Haralick [15] export some of the statistical parameters that describe the texture characteristics, Then 13 features (step d = 1, the angle of 0 degree direction) are extracted : Energy, Correlation, Inertia, Entropy, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Difference Average, Difference Variance, Difference Entropy, Information measure of correlation , Information measure of correlation. The common used definitions are shown as follows:

Energy:

$$ASM = \sum_{i=1}^{l}\sum_{j=1}^{l}P(i,j)^2 \qquad (1)$$

Energy mainly reflects texture thickness. The larger the value, the more coarse texture . The maximum of 1, p (i, j) is only one element in the value of 1, other is 0. The minimum is 1 / L. uniform distribution of p (i, j), image gray difference.

Entropy:

$$\text{Ent} = -\sum_{i=1}^{l}\sum_{j=1}^{l} P(i,j) lb\{P(i,j)\} \qquad (2)$$

Entropy reflects texture complexity. The greater the entropy value, texture is more complex; The maximum is lb (L^2). p (i, j) is evenly distributed. Smaller entropy value, texture, more simple, the minimum is 0, p (i, j) only one element value is 1, the other is 0.

Inverse Difference Moment:

$$\text{IDM} = \sum_{i=1}^{l}\sum_{j=1}^{l} P(i,j)/(1+(i-j)^2) \qquad (3)$$

Inverse Difference Moment local image gray balance metric, A maximum of 1, p (i, j) are located in the main diagonal, the local image gray unchanged, A minimum of 1 / (1 + (L-1)^2), p (i, j) only distributed in the (L 1) or (1, L) point, the whole image and gray-scale difference.

Inertia:

$$\text{Con} = \sum_{i=1}^{l}\sum_{j=1}^{l} (i-j)^2 P(i,j) \qquad (4)$$

It reflects the image intensity and texture. The larger the value, texture groove deeper, the visual effect was clear. The maximum (L -1), p (i, j) are only distributed in the (L 1) or (1, L) point. Minimum of 0 and P (i, j) are located on the main diagonal, the local image completely uniform, no gray-scale difference, no contrast.
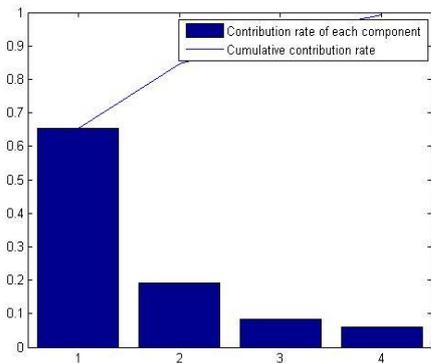
Figure 4.   The contribution rate of the first group of experiments

In order to find feature vector with the crowd density high correlation, we use PCA in the13 feature vector for data dimension reduction, and strong correlations of four dimensional feature vector is obtained. Figures (4-7) are the contribution rate of the 4-dimensional feature vector dimensionality reduction for each group of experimental images respectively; they are accounted for more than 99%.
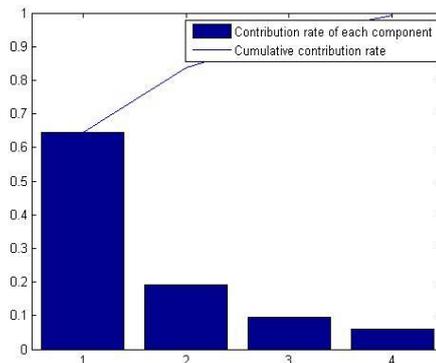
Figure 5.   The contribution rate of the second group of experiments
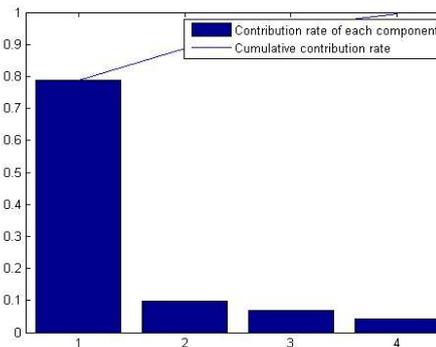
Figure 6.   The contribution rate of the third group of experiments
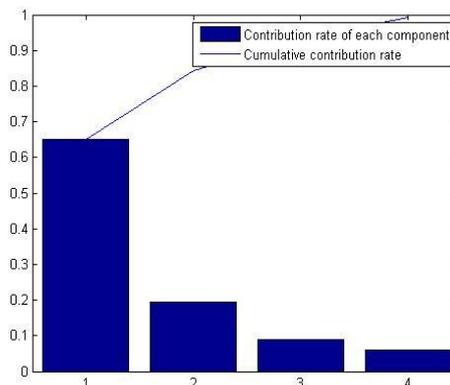
Figure 7.   The contribution rate of the forth group of experiments

C. Tamura Texture Feature Extraction

In the psychological research Based on human visual perception, Tamura et al. [16] proposed the expression of texture feature. Tamura texture characteristics of six components corresponding to the psychology Angle on the texture characteristics of six kinds of properties. They are coarseness contrast directionality line likeness regularity and roughness. When the crowd density is high, coarseness, contrast, direction, roughness of Tamura texture features is smaller. On the other hand, the population density is small, the value of Tamura texture features is high. The data of some experiment images as follows:

TABLEI.   THE EXTRACTION RESULTS OF EXPERIMENTAL IMAGE TAMURA TEXTURE FEATURES

Tamura texture feature	Coarseness	contrast	direction	Rough-hness
Image1	56.9190	46.7000	1.0884	101.889
Image2	56.8888	46.0192	1.0753	101.179
Image3	56.8801	45.3414	1.0207	100.562
Image4	56.8708	45.0668	1.0169	100.345
Image5	56.8671	44.6839	1.0069	101.003
Image6	56.8580	44.3157	0.9727	99.8334
Image7	56.8198	43.6214	0.9870	99.6587
Image8	56.8187	43.3374	0.9959	99.4735
Image9	56.7711	42.9091	0.9509	94.3515
Image10	56.7367	42.7787	0.9397	94.2932

The experiment images are from a video of the crowd. The crowd flow is from little to much and then gradually reduce after the monitoring scope. So, the density of crowd is small at beginning, and then it becomes high gradually. From table I, we can know that the value of Tamura texture feature become small when the density is high. (there is a little exception when calculate the value of direction for some reasons.)

The formula of Coarseness, contrast, direction, roughness is as follows:
Coarseness:

Calculating moving averages, for the window 2k*2k, the moving average are:

$$a_k(i,j) = \sum_{i'=i-2^{k-1}}^{i+2^{k-1}-1} \sum_{j'=j-2^{k-1}}^{j+2^{k-1}-1} \frac{p(i',j')}{2^{2k}} \qquad (5)$$

Calculate the horizontal and vertical deviation:

$$c_k(i,j) = \max(|a_k(i-2^{k-1},j) - a_k(i+2^{k-1},j)|,$$
$$|a_k(i,j-2^{k-1}) - a_k(i,j+2^{k-1})|) \qquad (6)$$

To determine the size of the window:

$$\hat{k}(i,j) = \underset{k}{\arg\max}\, c_k(i,j) \qquad (7)$$

Calculate the average size of the window and get the

$$Fcrs = \frac{1}{wh} \sum_{(i,j)} 2^{\hat{k}(i,j)} \qquad (8)$$

Contrast:

$$Fcon = \frac{\sigma}{\sqrt[4]{\alpha_4}} \qquad (9)$$

Direction:

$$Fdir = \sum_{p}^{n_p} \sum_{\theta \in w_p} (\theta - \theta_p)^2 H_D(\theta) \qquad (10)$$

Roughness is the synthesis of the Coarseness and contrast two properties.

D. SVM Classifier Prediction

LIBSVM is a library for Support Vector Machines (SVMs). The LIBSVM is developed and designed by Lin Chih-Jen Taiwan University. It is a simple and easy-to-use support vector machines tool for classification (C-SVC, nu-SVC), regression (epsilon-SVR, nu-SVR), and distribution estimation. The package can be obtained on the Internet for free [17]. The goal is to help users to easily apply SVM to their applications. LIBSVM has gained wide popularity in machine learning and many other areas.

In this part, we fused four texture features extracted by Tamura and GLCM. Crowd density leveled by manual calibration. Qualitative description the crowd density monitoring area, it is classified according to crowd density. Crowd density is usually divided into five types. They are very low, low, medium, high, and very high. The scope of each type is related to the service level. Table II shows the Polus about the definition of service level in 1983. Table III shows the relationship between density of classification and the service level (monitoring area is 15 m^2). Due to the camera is stationary in this the experimental data. Population flow is from little to much and then gradually reduce after the monitoring scope. According to the number of features in the data sets, this paper will divide into the crowd density four parts: low, low, medium and high, which are labeled 1, 2, 3, 4 respectively.

TABLEII.   THE SERVICE LEVEL DEFINITION

The service level	Crowd density (person/m^2)
A:Normal people	≤0.6
B:Dense stream of people	0.6~0.75
C1:Dense stream of people	0.75~1.25
C2:Very dense stream of people	1.25~2.0
D:Congestion crowd	>2.0

TABLEIII.   RELATIONSHIP BETWEEN DENSITY CLASSIFITON AND THE SERVICE LEVEL

Density classification	The number range	Density level	service level
Very low	0~10	1	A, B
low	11~20	2	C1
medium	21~35	3	C2, D
high	36~50	4	D

TABLEIV.   BENCHMARK DATA

Data set (PETS2009)	Total images	training set	test set
S1_L2.14-06(1)	200	160	40
PETS2009 S1_L2.14-31(1)	130	90	40
PETS2009 S1_L1.13-57(1)	220	--	40
PETS2009 S1_L1.13-59(1)	240	200	40

III.   EXPERIMENTAL RESULTS

All the images used in this paper are derived from PETS2009 benchmark data [18]. Table IV shows that the training set and test set of every experiment group. The total number of images is 790 in our experiment.

The four groups of images were adopted from the same camera in the same scene but at different times and under different lighting conditions. Images are JPG format of the size 768 * 576. Final experiments classification prediction results as shown in Figure (8,9,10,11).
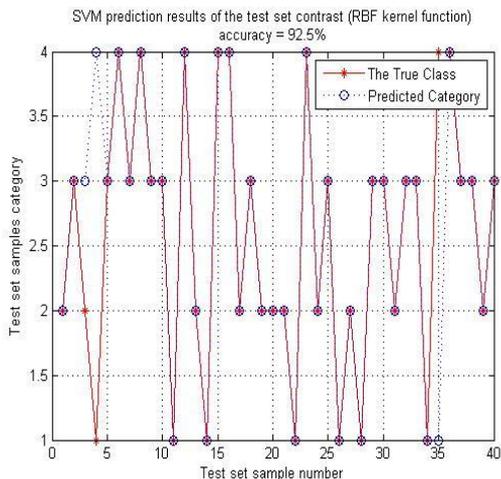
Figure 8.   The predicted results of the first set of data
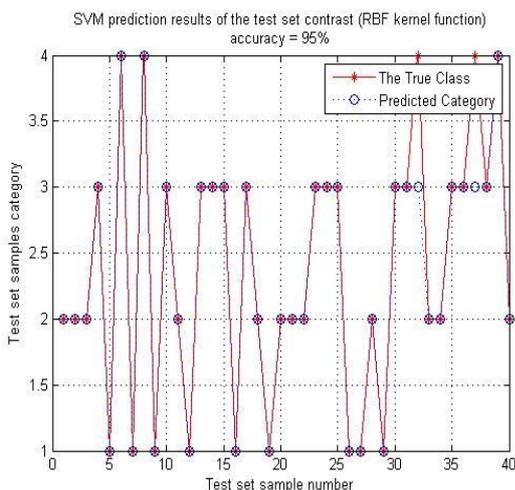
Figure 9.   The predicted results of the second set of data
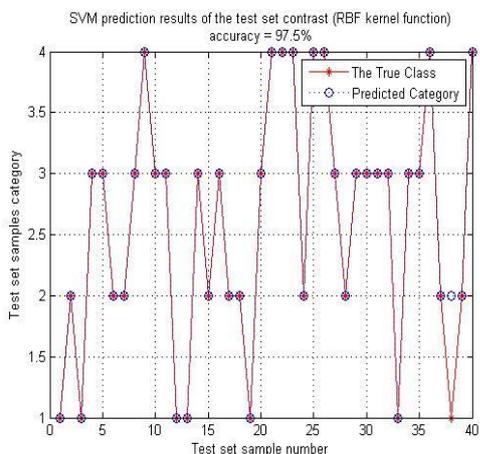
Figure 10. The predicted results of the third set of data

The experimental groups are shown in table IV and V. The numbers of first group are 200 images. we selected 160 images randomly for training and the remaining 40 to test. The numbers of second group are 130 images and then selected 90 images randomly for training and the remaining 40 to test. The numbers of third group are 220 images and we selected 180 images randomly for training

and the remaining 40 to test. The numbers of forth group are 240 images. we selected 200 images randomly for training and the remaining 40 to the test. If we extract the texture feature only with GLCM to estimate the density of crowd, the average accuracy rate is 90%. Table VI is the result compared with the method which just use gray level co-occurrence matrix (GLCM) to extract the texture features. At the same time, the comparison with other methods by the similar way is shown in table VII [19]. From table VII We can know that Marana and Wu extracted the same texture feature but the feature pretreatment are different. Chan used Pixel and texture feature to estimate the crowd density. Our method extracted texture feature by GLCM and Tamura and the feature pretreatment was PCA. Finally, we can see that the accuracy rate of our method is more exact than others.
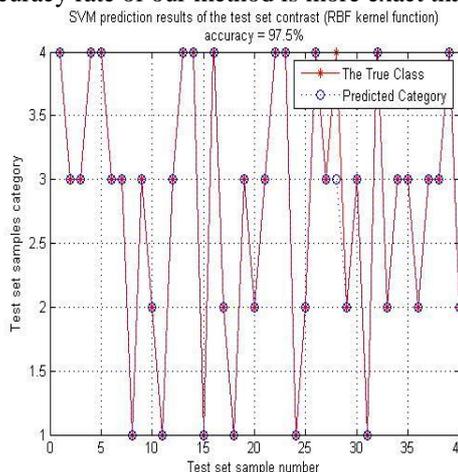
Figure 11. The predicted results of the fourth set of data

TABLEV. PREDICTION ACCURACY STATICS TRAINING SAMPLE AND TEST SAMPLE

Experimental group	Accuracy of training samples	Accuracy of test samples
The first group(S1_L2.14-06)	99.375%(159/160)	92.5% (37/40)
The second group(S1_L2.14-31)	97.778% (88/90)	95% (38/40)
The third group(S1_L1.13-57)	99.444% (179/180)	97.5% (39/40)
The fourth group(S1_L1.13-59)	99.5% (199/200)	97.5% (39/40)
average accuracy rate	99.024%	95.5%

IV.   CONCULSION

Considering the real-time and accuracy of the estimation the crowd density, extracting feature vector which has high correlation with the crowd density is worth studying. This paper introduces a new image texture feature called Tamura for crowd density estimation. The features GLCM are reduced dimension by PCA method, which can obtain the features with high correlation to the crowd scenes. The experimental results show that the method can improve the classification accuracy of prediction. However, the calculation of support vector machine (SVM) method is large, and how

to obtain the optimal parameters which will be further studied in the future.

TABLE VI.   COMPARED WITH THE METHOD ONLY USE GLCM TO EXTRACT TESTURE FEATURE

Experimental group	Accuracy of training samples	Accuracy of test samples
(S1_L2.14-06)	90%(36/40)	92.5%(37/40)
(S1_L2.14-31)	87.5% (35/40)	95% (38/40)
(S1_L1.13-57)	90% (36/40)	97.5%(39/40)
(S1_L1.13-59)	92.5% (37/40)	97.5%(39/40)
average accuracy rate	90%	95.5%

TABLE VII.   COMPARED WITH OTHER EXPERIMIMENTAL RESULTS

Author	People feature	Feature pretreatment	crowd analysis	Accuracy
Marana[3]	GLCM	--	neural net	High-density: 93.%
Wu[8]	GLCM	Projection deformity correction	SVM	91.3%
Chan[9]	Pixel and Texture	Projection deformity correction	Statistical-regression	more than 90%
Our	GLCM and T - amura	PCA	SVM	95.5%

ACKNOWLEDGEMENT

REFERENCES

[1] Nikos Paragios, Visvanathan Ramesh. A MRF-based Approach for Real-time Subway Monitoring. IEEE Trans, 2001(I) pp. 1034-1040.

[2] H. Rahmalan, M. S. Nixon, J.N. Carter. On Crowd density estimation for surveillance. *IEEE Int Conf. Crime Detection and Precention*, 2006 pp. 540-545.

[3] Wu Xinyu, Liang Guoyuan, Lee Ka Keung, Xu Yangsheng. Crowd Density Estimation Using Texture Analysis and Learning. *IEEE International Conference on Robotics and Biomimetics*. 2006 pp. 214-219

[4] Chan A B, Liang Z S J, Vasconcelos N. Privacy preserving crowd monitoring: Counting people without people models or tracking. 2008

[5] Qing WEN, Chengcheng JIA et al. People Number Estimation in the Crowded Scenes Using Texture Analysis Based on Gabor Filter. *Journal of Computational Information Systems* 7: 11 (2011) pp. 3754-3763

[6] Zhaoxiang Zhang, Min Li. Crowd density estimation based on statistical analysis of local intra-crowd motions for public area surveillance. *Optical Engineering*. 51(4) , 2012

[7] D. B. Yang, H. H. Gonzalez-Banos, L. J. Guibas. Counting people in crowds with a real-time network of simple image sensors. Proceedings. *Ninth IEEE International Conference on Computer Vision*, 2003, 1 pp. 122-129

[8] Antonio Albiol, Maria Julia Silla, Alberto Albiol and Jos é Manuel Mossi: Video Analysis using Corner Motion

Statistics. *Proceedings 11th IEEE International Workshop on PETS*, Miami, June 25, 2009.

[9] Donatello Conte, Pasquale Foggia, Gennaro Percannella et al, A Method for Counting Moving People in Video Surveillance Videos, *EURASIP Journal on Advances in Signal Processing Volume* 2010.

[10] Shafin Rahman, Sheikh Motahar Naim et.al. Performance of PCA Based Semi-supervised Learning in Face Recognition Using MPEG-7 Edge Histogram Descriptor. *Journal of Multimedia*. 2011 (10).

[11] Shaobo Zhong, Dongsheng Zou. Web Page Classification using an Ensemble of Support Vector Machine Classifiers. *Journal of Networks*. 2011 (11).

[12] [Haiyu Song, Xiongfei Li et.al. Adaptive Feature Selection and Extraction Approaches for Image Retrieval based on Region. *Journal of Multimedia*. 2010 (2).

[13] Gang Xie, Tianrui Cao, Chengdong Yan. Texture Features Extraction of Chest HRCT Images Based on Granular Computing. *Journal of Multimedia*. 2010 (12).

[14] Chengcheng Gao, Xiaowei Hui. GLCM-Based Texture Feature Extraction. *Journal of Computer Applications*. 2010.

[15] Haralick R M. Statistical and Structural Approaches to Texture. *Proc. of the IEEE*, V01. 67 No. 5 May 1979 pp. 786-804.

[16] Ideyuki Tamura, et al. Texture Features Corresponding to Visual Perception. *IEEE Trans. on Systems, Man and Cybernetics*, 1978, 8(6) pp. 460 - 473.

[17] http://www.csie.ntu.edu.tw/~cjlin/

[18] http://www.cvg.rdg.ac.uk/PETS2009/

[19] Su Hang, Zheng Shibao et al. Survey of Crowd Flow and Density Estimation in Video Surveillance. *Video Engineering*. 2009, 33(11)

Bobo Wang, male, was born in Hebei, China, Oct 1986,received his Bachelor of Engineering degree in Computer Science and Technology from North College of Beijing University of Chemical Technology, China, in 2011.He is studying as a graduate student in Beijing Union University. His research interests concentrate on image processing and distributed system.

Hong Bao, male, was born in Beijing, China, received his Ph.D. degree in computer science from school of computer and information technology from Beijing Jiaotong University Beijing, China and M.S. degree in information technology from Liverpool University, U.K.. He is now a professor of Beijing Union University, Beijing, China. His current research interests include image processing, visual perception and distributed system.

Shan Yang, female, was born in Hunan, China, Nov. 1989, received her Bachelor of Engineering degree in Communication Engineering from Beijing Union University, China. She is studying as a graduate student in Beijing Union University now. Her research interests concentrate on image processing and distributed system

Haitao Lou, male, was born in Beijing, China, 1975, received his master degree from Beijing University of Posts and Telecommunications, China. His main research directions are digital image processing and artificial intelligence.

Interaction Design of National Dance Based on Realistic 3D Character

Guoxin Tan

National Research Center of Cultural Industries, Central China Normal University, Wuhan, China
Email: gxtan@mail.ccnu.edu.cn

*Corresponding author Chuanming Sun

National Engineering Research Center for e-Learning, Central China Normal University, Wuhan, China
Email: shuncm@163.com

Ping Wang

Turku School of Economics, University of Turku, Turku, Finland
Email: applepiewp@hotmail.com

Abstract—**Great improvement has been achieved in the protection of national dance through multimedia technology. An interactive design approach for national dance based on realistic 3D character is proposed in this paper. This approach comes with three sub steps: Firstly, the realistic face was reconstructed based on a front photo, in this step, the facial feature points of photo are selected interactively; and then, the realistic face was built through texture mapping and fusion based on the standardized face model database. Secondly, the construction of realistic body is realized through introducing the stretching model and the vector differential adjuster. Finally, the interactive display of national dance is realized by virtual reality engine. The experimental results show that this approach is user-friendly, and can generate a high-quality realistic 3D character in real time while protects the worthy cultural heritage effectively at the same time.**

Index Terms—**Realistic 3D Character, National Dance, Interaction Design**

I. Introduction

National dance is an important part of intangible cultural heritage. With the great advancement of economy, many precious national dances are struggling to survive due to the strike of modern civilization. One of the most commonly used approaches to protect national dance veritably and systematically are taking photos, recording videos and other multimedia tools. The problem is: the recording outlooks for dancers' motion are limited, as a result, the reusability and editability is poor.

With the development of computer software and multimedia technology, it has become a trend to protect national dance by three-dimensional digital technology. The current achievements have several shortcomings, the actors in it lack of realistic and insufficient interaction have affect the user immersion and initiative seriously [1]. As a result, it's important to design professional characters and apparels to provide a more intuitive form for digital dance according to the special characteristics of dancing. However, the generate process of realistic character is often quite complex, and it's not easy for general users to obtain the necessary equipment and corresponding databases. In the field of computer graphics, realistic characters are the avatar of users in a virtual environment, which can be widely used in human-computer interaction, virtual simulation, and other fields. For example, the participants can be "embedded" in a virtual environment in the production of digital national dancing, making them a form of existence in the virtual world, through which the interest and enthusiasm of the participants could be greatly aroused.

There are four dominant approaches to build realistic characters model currently, they are as follows:

Firstly, using professional modeling software (such as: Autodesk Maya, 3Ds Max), which can create a human model by operate the most basic modeling elements interactively. Although this approach, users have the greatest authority to control, high level of creativity and sufficient familiarity are needed to perform well in the modeling software.

Secondly, using professional capture devices like the three-dimensional scanner to collect data, such as Devernay [2] and Bickel et al [3]. The advantage is that the accuracy of the final model is relatively high, but the results tend to have holes in the model, What's more, the operation and maintenance of these devices need professional training.

Thirdly, researchers like Sloan [4] taking interpolation operation on the vertices in the same location using a plurality of the samples, to obtain a new model. But the establishment of such a database spends a lot of time and resources; what's more, it does not take the generation of the entire human body into consideration.

The forth approaches comes that Hilton [5] and Liao et al [6] use the shape features which are extracted from the individual photos to modify the generic model. But
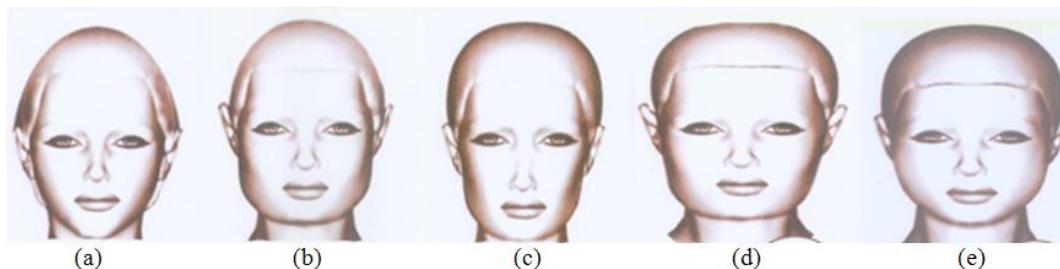
Figure 1.   The types of face, (a) is oval-shaped type; (b) is inverted triangle type; (c) is long-shaped type; (d) is square type; (e) is round type.

Hilton's approach is likely to cause the model cracks. Liao's approach can generate realistic head model through only a single photo, but it need to select many characteristic points.

This research absorbed the advantages in the third and fourth approaches, and proposed a unified approach which can generate realistic 3D faces based on a single font photo and generate realistic 3D body based on the vector difference. Then the realistic character will be bond with the motion capture data of national dance, and then be displayed in the multimedia platform interactively. There are several sub-steps in the face modeling. Firstly, the standardized face model database should be constructed, and interactively select the facial feature points of the input picture. After that, we'll choose the best matched face model based on the face features. Secondly, the approach of triangular deformation and bilinear interpolation are used to achieve texture mapping from photo to three-dimensional model, and the introduction of Alpha picture of face to realize the fusion transition from the photo texture to the neutral texture of the model in order to generate realistic 3D face. In the step of body generation, firstly, a reference model, a vertical stretching model and a horizontal stretching model are constructed in accordance with the model structure of the human body; secondly, a vector difference adjuster is employed to interactively adjust the control points in the reference model; lastly, a realistic character model is automatically generated based on the control points' coordinates. This approach can overcome the problems associated with other approaches, such as: the need to master professional model software, cracks or holes generated in the result model, and make it easy for users to interactively build a realistic character model. Through the application of this approach into national dance, we can promote the digital preservation and propagation of national dance.

II.   REALISTIC FACE GENERATION

The realistic sense of three-dimensional face can be reflected in the perspectives of facial model and texture. In the approach proposed in this paper, the users will participate in the texture specification and model adjustment. Both of these two processes are highly simplified, and we can view the results in real-time. Five steps in detail are contained in the approach proposed in this paper:

Step1. Feature point calibration: Get the front face photo, and mark 11 feature points which are spread over eyes, nose, mouth and so on to determine the approximate location of the five sense organs and facial form.

Step2. Model matching: Construct the standardized three-dimensional face model database, and match the most similar model in the corresponding face database to minimize the differences of facial bones.

Step3. Texture mapping: Map the texture feature of photo face on the matching model through the algorithm of triangular deformation and bilinear interpolation.

Step4. Skin color fusion: Use an Alpha picture to achieve the gradient from the overlay texture to model texture in the face boundary area.

A. Feature Point Calibration and Model Matching

If the face photo is obtained, a suitable face model should be selected according to the positional relationship of the facial feature points. Now some standardized 3D face model database had already been constructed, such as UND [7], BU-3DFE [8], and BJUT [9]. These face model databases had been widely used in face generation, face recognition and other fields.

Different 3D face model database have different standards on race, age, gender, light, etc. The approach of this paper is mainly for the Chinese people. We built a Chinese face model database with three-dimensional scanning equipment. The types of face were divided into five types which is show in Fig. 1, such as oval-shaped, inverted triangle, long-shaped, square and round. However, the neutral head model (as shown in Fig. 2)is completed by the reference of CANDIDE-3 neutral face model [10].

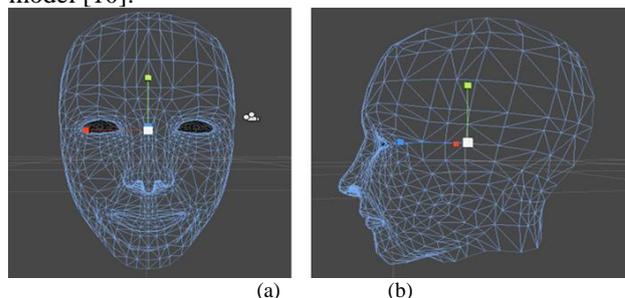Figure 2.   The neutral head model, (a) is the front view and (b) is the side view.

After the completion of the database, the model must get a pretreatment, including repairing model crack, binding neutral texture, size normalization, calibrating models feature points and so on. MPEG4 standard specifies 84 feature points in the neutral face [11]. According to the requirement of face model matching and

texture mapping, in this paper, we select 11 feature points on the input photo and the face model, as shown in Fig. 3. Through which, the facial features and facial type can be confirmed approximately, so that we can get the most suitable model.

Model matching approach is as follows: First of all, number the feature points to calculate the distance and proportionate relationship between the main feature points; Secondly, use literature [12] for reference to determine the general facial type; Finally, match the Euclidean distance of the corresponding feature point between model and photo in the corresponding facial type database, and get the nearest 3D model.

$$E = \sum_{i=1}^{n} (D_i - D_i\,')^2 * \lambda_i \qquad (1)$$

And D_i is the distance between two feature points in the photo, $D_i\,'$ is the corresponding distance in 3D model, and λ_i is the weight value for each feature point distance, n is the quantity of the corresponding distances.
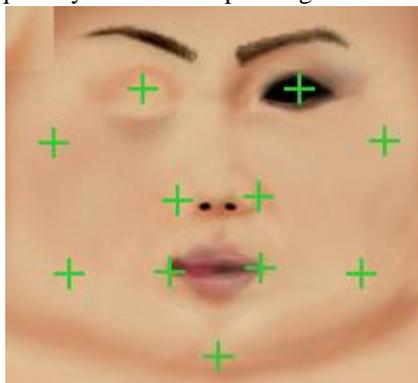
Figure 3.   Feature points calibration.

B. Texture Mapping

High-quality texture mapping can make the model more realistic. In the approach proposed in this paper, the face texture generated consists of two layers texture: the first layer is the model neutral texture, it is a default texture which is blinded on the model; the second layer is the cover texture, it should be obtained from a good quality photo. The feature points' position marked by neutral texture and coverage texture would form two 2D point sets—Set V_1 and Set V_2. Due to the discrepancy existed in hand-marking, it's much more likely that the corresponding points in V_1 and V_2 are not in the same location precisely. So a mapping approach is needed to combine the two sets naturally. Take $P_1 \in V_1, P_2 \in V_2$, and extend P_1, P_2 to the 3D vector, $P_1 = (X_1, Y_1, 1), P_2 = (X_2, Y_2, 1)$. After obtaining the transformation matrix for each point, the transformation of feature points needs to be applied to the pixels around the feature point. The triangular deformation in the process will be used to apply the cover texture to the model texture. The pixels in the triangle are affected by the vertexes of triangle. Assuming that the three vertexes

of the triangle are P_1, P_2, P_3. Set these vertexes as the three columns of a matrix, forming a 3 * 3 matrix (P_1, P_2, P_3). Assume that the transformed triangle is $(P_1\,', P_2\,', P_3\,')$, then:

$$M * (P_1, P_2, P_3) = (P_1\,', P_2\,', P_3\,') \qquad (2)$$

For the Point P_i in the triangle, $P_i^{'}$ will be derived by $M * P_i$, then, the mapped coordinates of all the points in the triangle can be obtained. For the unmapped point, all the coordinate values are integers, the coordinate value after mapping may not be an integer, so the pixels are unable to be mapped one-to-one with texture picture. Therefore, a sampling approach is needed for taking sample for texture color. In this paper, bilinear interpolation approach (as shown in Fig. 4) which can provide a good quality effect is used for sampling.
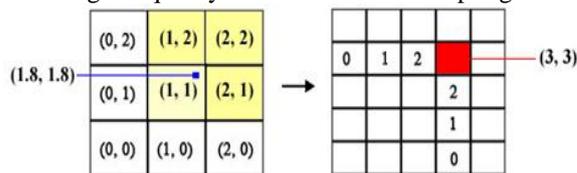
Figure 4.   Bilinear interpolation approach.

C. Skin Color Fusion

After texture mapping, color difference may appear obviously in the texture boundary. This is because the color difference between the neutral Fig and overlay Fig. There are a variety of approaches to eliminate this difference, Rocchine [13] achieved the fusion transition by the linear interpolation on texture boundary, but it is difficult to achieve a smooth transition if the texture color difference is too large or texture overlapping ranges is too small. The approach used in this part is to draw an Alpha Fig with natural transition effects, and eliminate the boundary mutation by setting the transition area of boundary.
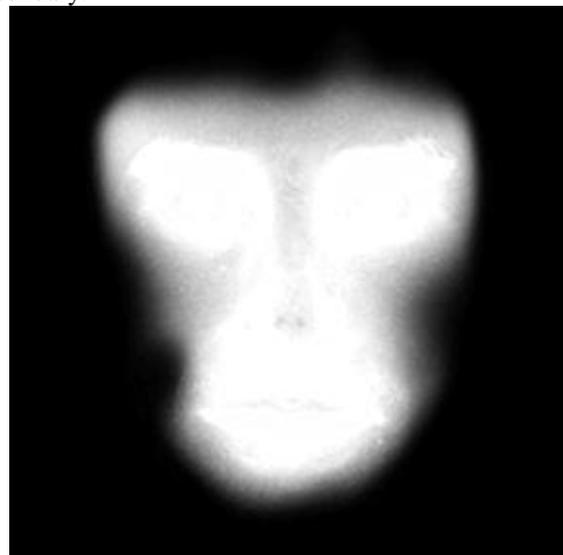
Figure 5.   The alpha Fig.

| (a) | (b) |
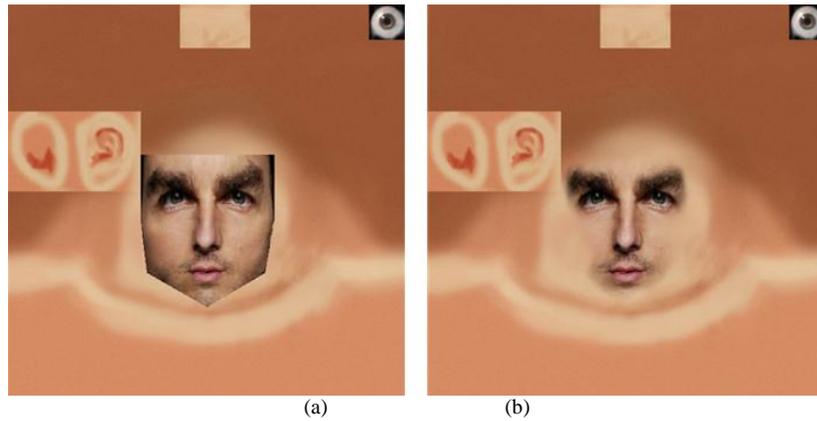
Figure 6.   The effect of skin color fusion, (a) is the effect without color fusion and (b) is the final fusion effect.

| (a) | (b) | (c) |
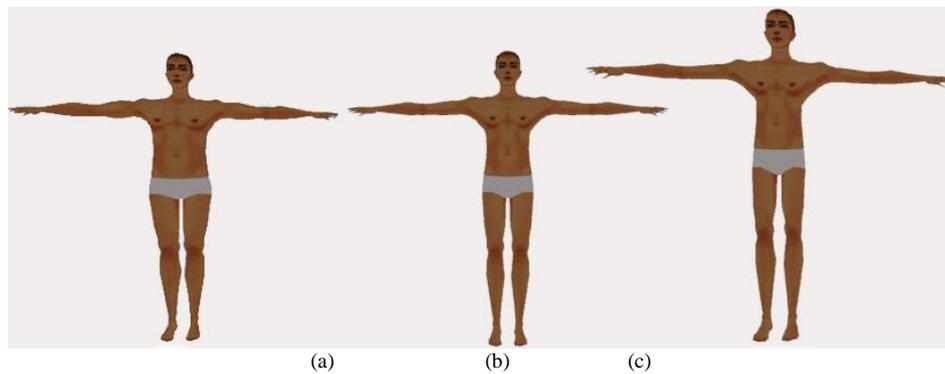
Figure 7.   Body model, (a) is horizontal extension model, (b) is reference model, (c) is vertical extension model.
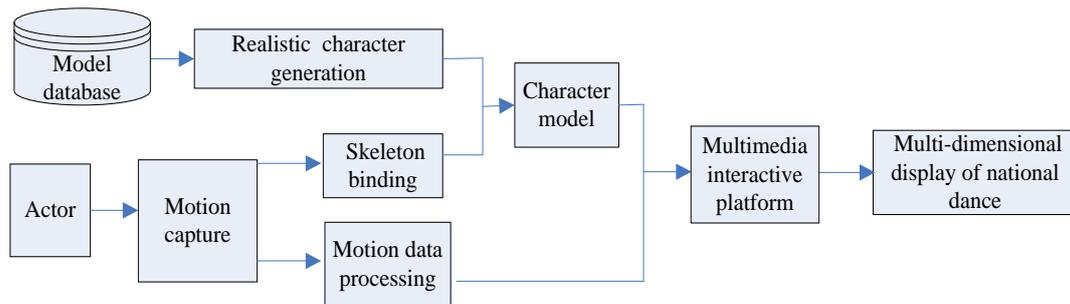
Figure 8.   The step of interactive platform build.

The Alpha Fig which is shown in Fig. 5 forms a transition area in the face boundary by the gradation of gray. In the fusion process, the neutral Fig, overlay Fig, and the Alpha Fig will be processed as the same size. The mixed formula of these three Figs is as follows:

$$C(x, y) = C_{neutral}(x, y) * (255 - C_{Alpha}(x, y)) \\ + C_{overlay}(x, y) * C_{Alpha}(x, y) \qquad (3)$$

and $0 \le C_{Alpha}(x, y) \le 255$

The pixels in Alpha Fig would participate in the mixture as the mixed interpolation weight, and the natural fusion would be finally achieved. The fusion effect is shown in Fig. 6.

III.   REALISTIC BODY GENERATION

It is still a very challenging task to build a realistic body model interactively. The approach in this part can

overcome the shortcomings existing in other approaches, for example: the requirement of mastering professional model software, cracks or holes generated in the model. This part makes it easy to build realistic virtual character models interactively. The steps in detail are as follows:

Step1. Design a set of realistic body models including a reference model, a vertical extension model and a horizontal extension model.

Step2. Obtain the vector differences between control points in the reference model and the horizontal and vertical stretching models.

Step3. Design a vector difference adjuster, and adjust every control point in the reference model interactively to generate realistic body model.

A. Reference Model

Most of the body height of Chinese adults varies from 150cm to 185cm [14, 15]. Take the body difference between male and female, we constructed a female and a

male standard reference model respectively. The height range of the reference model was set between 155CM and 185CM, and was divided into seven segments with 5CM per segment. So, a single model height variation will be in the range of -2.5cm-2.5cm. The other two models with the same structure of reference model are also designed: a vertical extension model which is higher than the reference model; a horizontal extension model which is fatter than the reference model. The three models are shown in Fig. 7.

B. Vector Differential Calculation

The key step of this approach is to obtain vector coordinates of control points rapidly and accurately. It can be solved by designing the model file designed into an m-tree data structure. The right and left hand, leg, and the chest of character model, which are marked above the m-tree, can be quickly found by traversing the m-tree, and then the vector coordinates of their control points are obtained. Using this approach, coordinates vector of the control points of the reference model, vertical extension and horizontal extension model are obtained respectively $\vec{M}_{Dp} = \{x, y, z\}$, where x, y, z respectively represent the coordinates of x, y, z axis, $M \in \{H, F, C\}$, $D \in \{W, L, O, N\}$, $p \in \{r, l\}$, H represents the hand, F represents the foot, C represents the chest, W represents horizontal extension model, L represents the vertical extension model, O represents the original model, N represents the new model, r represents right side, and l represents left side.

After obtaining the vectors of the three models, the following formula is used to calculate vector differences of horizontal extension model and reference model, vertical extension model and reference model:

$$\Delta \vec{M}_{D'p} = \vec{M}_{D'p} - \vec{M}_{D''p}$$

and

$$D' \in \{W, L\}, D'' = 0 \tag{4}$$

C. The Design of Vector Differential Adjuster

In this part, a vector differential adjuster is designed to allow users modify model shape in real-time. The model is divided into different components, such as head, hands, feet and chest. Except for the head, two kinds of adjuster are designed for the other three components. One controls the parameter of vertical increment, and the other controls the parameter of horizontal increment.

After obtaining every vector differential above, adjusters are employed to control the incremental impact for the various components of the reference model in this part. The control points' vector coordinators of the new model are obtained through continuously adjusting:

$$\vec{M}_{Np} = \vec{M}_{Op} + \alpha \Delta \vec{M}_{Wp} + \beta \Delta \vec{M}_{Lp} \tag{5}$$

While α represents the incremental parameters of horizontal extension model, β represents the incremental parameters of vertical extension model.

When the body is extended or shortened, the joints between body and foot would have a crack. In order to solve this bug, the formula (6) and (7) are used on feet parts [16]:

$$\vec{F}'_{Nl} = \vec{F}_{Ol} + \alpha \Delta \vec{F}_{Wl} + \beta_f \Delta \vec{F}_{Ll} + \beta_c \Delta \vec{C}_L \tag{6}$$

$$\vec{F}'_{Nr} = \vec{F}_{Or} + \alpha \Delta \vec{F}_{Wr} + \beta_f \Delta \vec{F}_{Lr} + \beta_c \Delta \vec{C}_L \tag{7}$$

While β_f represents the feet's control parameters of vertical extension model, β_c represents the chest's control parameters of vertical extension model.

IV. NATIONAL DANCE INTERACTION DISPLAYING

It is only a technical way to design a realistic role and bind the dance motion which is captured by motion capture system. It is only an important step to realize the digitized recovery of dance, but not our ultimate goal. Only through a convenient way to display, even with some entertainment features, and in the aid of high-speed network, the masses can have the opportunity to watch and understand this recovery achievement [17]. It can promote the protection and spread of the cultural heritage effetely. In this paper, Funeral dance of Tujia nationality, a Chinese national dance, is selected as a typical demonstration. Through recovering the realistic actor, the national dance is displayed by multi-dimensional based on the multimedia interactive platform, which realized the digital representation and spread of national dance [18]. The interaction design process is shown in Fig. 8. Through the module of behavioral interaction, the user can manipulate the behavior of realistic actor and watch the activities of the recovered group culture interactively.

V. EXPERIMENTAL AND ANALYSIS

In order to verify the robustness and accuracy of the proposed approach, the experimental system is developed based on the Visual Studio 2010 platform. The configuration of experimental computer is Inter i5-2400 CPU@3.10GHz and 4GB memory. In this experiment, the effectiveness will be certificated from the two aspects: the realistic character generation and multimedia interactive display.

A. The Generation Experiment of Realistic Character

The 3D face model database is constructed by capturing the different user face with the MH/MHT 3D scanner, and the models are standardized pretreatment. The face type of input photo is determined through the proposed approach. Then the most similar model is matched in corresponding face database, and achieving the texture mapping and fusion automatically. According to users' height and body proportions, they can adjust the combination deformation for the different parts of the model and even the whole body in vertical and horizontal direction. Fig. 9 is a 3D head model and the body model which was obtained by adjusting the composition changes. Therefore, the approach in this paper can be used to control the model deformation interactively, and to generate the realistic character quickly.

Figure 9.   The final effect of realistic character.

Figure 10. The final effect of funeral dance. B. Multimedia Interactive Display

In this part, the dance motion is obtained from live performance based on the principle of true representation. Therefore, the experiment system collects the dance motion from dance actor by motion capture system. Then, the funeral dance is reappeared dynamically and realistically by binding the realistic character.

Finally, based on the archaeological map, recovering the real 3D scene of Badong County, building the performing venues, and expressing the funeral dance visually by human-computer interaction, Fig. 10 is the final effect of this dance.

C. Result Analysis

For the analysis of realistic character generation, this approach is a simple, yet very effective way to interact with users and has the most similarity with the photo. This approach provides the function of skin color fusion, and can adjust the model to achieve the optimal model, and less feature point were selected which reduced the complexity of the algorithm and the computation time.

For the application of national dance display, the advantages come as follows:

(1)Through the interactive system based on realistic character, visitors' immersion are strengthened and learning desire stimulated, which can guide the user to understand the relative display content better.

(2)Taking advantage of this platform, users can realize the interaction and learning of national dance in a friendly interface in a very simple way.

VI.   CONCLUSION

In this paper, an approach of national dance protection based on realistic character is proposed. It provides a more realistic interactive display platform for the field of dancing. As a tool, it is a symbol that the traditional dancing art can get faster and better development; while at the same time, as an idea, it represents a big progress in the combination of dance and digitization. In addition, the approach of realistic character generation are of better robustness which does not require users to master professional modeling software, and it's convenient for the user to build realistic character models. It can also be widely used in the virtual campus, virtual libraries, and virtual museum and so on.

REFERENCES

[1] Jianyi Liu, Kuizhi Mei, Fangfang Wang. Interactive Poisson Photometric Propagation for Facial Composite. *Journal of Multimedia*, Vol. 8, No. 2, pp.153-160, 2013

[2] Devernay F, Faugeras OD. Computing differential properties of 3-D shapes from stereoscopic images without 3-Dmodels. In Proc. Of Computer Vision and Pattern Recognition, pp. 208-213, 1994

[3] Bickel B, Botsch M, Angst R, etal. Multi-scale capture of facial geometry and motion. ACM Transactions on Graphics, Vol. 26,No.3, pp. 30.1-30.10,2007

[4] Sloan P, Rose C, Cohen M. Shape by example. In Proceedings of the 2001 symposium on Interactive 3D graphics. *ACM Press New York,* NY, USA, pp. 135-143, 2001

[5] Hilton A, Beresford D, Gentils T, et al. Virtual people: capturing human models to populate virtual worlds. Computer Animation, pp. l74-185, 1999

[6] Hai-bin Liao, Qing-hu Chen, et al. Rapid 3D face reconstruction by fusion of SFS and Local Morphable Model. *Visual Communication & Image Representation,* Vol. 23, No. 8, pp. 924-930, 2012

[7] Chang KI, Bowyer KW, Flynn PJ. An evaluation of multimodal 2D+3D face biometrics. PAMI 2005, 27(4), pp. 19-24

[8] Sihai Zheng, Layuan Li, Yong Li, A QoS Routing Protocol for Mobile Ad Hoc Networks Based on Multipath, *Journal of Networks,* Vol. 7, No. 4, pp. 691-698, 2012

[9] Jun Chen, Yuesheng Gu, Yanpei Liu, Grid Service Concurrency Control Protocol, *Journal of Networks*, Vol.7, No. 4, pp. 707-714, 2012

[10] Jörgen Ahlberg, CANDIDE-3-an updated parameterised face. Image Coding Group Dept. of Electrical Engineering, Linköping University, SWEDEN 2001.

[11] Jörn Ostermann, Animation of Synthetic Faces in MPEG-4. CA1998, pp. 49-55.

[12] XU Yanqiu, QIU Jianxin, Measurement and Classification of Shanghai Female Face Shape Based on 3D Image Feature. Journal of Shanghai University of Engineering Science, 25(1), pp. 60-64, 2011

[13] Min Zhao, Tao Zhang, Fangbin Ge, Zhijian Yuan, RobotDroid: A   Lightweight   Malware   Detection

Framework On Smartphones, *Journal of Networks*,Vol.7, No. 4, pp. 715-722, 2012

[14] Jing Ma, Shuli Sun, "Centralized Fusion Estimators for Multi-sensor Systems with Multiplicative Noises and Missing Measurements", *Journal of Networks,* pp. 1538-1545, Vol. 7, No. 10, 2012

[15] Yiran Wang, Bo Zhang, Shujian Li, "Differential Coherent Algorithm for Interferometric Acquisition of GNSS-R Software Receiver", *Journal of Networks*, Vol. 7, No. 10, pp. 1687-1695, 2012

[16] GuoxinTan, Tanghai Liu. Realistic Character Modeling Based on Vector Difference. *5th International Conference on Computer Sciences and Convergence Information Technology* 2010, pp. 192-197.

[17] Yanjun Long, Jianquan Ouyang. Research on Multicast Reliability in Distributed Virtual Environment. *Journal of Networks*, Vol. 8, No. 5, pp. 1130-1137, 2013

[18] Grant Schindler, Frank Dellaert. 4D Cities: Analyzing, Visualizing, and Interacting with Historical Urban Photo Collections. *Journal of Multimedia,* Vol. 7, No. 2, pp. 124-131, 2012

Guoxin Tan is a professor and doctoral tutor, vice dean of National Research Center of Cultural Industries. He is a member of Multimedia Technology Specialized Committee of China Computer Federation(CCF TCMT) . His current research interests include computer vision, digital protection of cultural heritage and multimedia information processing.

Chuanming Sun is a doctoral candidate in National Engineering Research Center for e-Learning. His current research interests include image processing, digital protection of intangible cultural heritage and human-computer interactive.

Ping Wang is a doctoral candidate in the subject of Information System in the University of Turku. Her current research interests include e-commerce, and digital tourism.

Image Segmentation Method Based on Snowfall Model

Yan YU, Jian-hua WANG *
Institute of Computer Science and Information Engineering, Harbin Normal University, Heilongjiang, Harbin, 150025, China
Email: Yuyan9999@vip.qq.com, WJH@vip.163.com

Madukasi Charles Nnaemeka
God's time Contracting Company, Federal Housing Estate, Onitsha, Anambra State, 430273, Nigeria
Email: cmakas4all@yahoo.com

***Abstract*—In order to improve accuracy of image segmentation, reduce the effect of noise on the cutting edge of image segmentation as much as possible, a new image segmentation method based on the model of the snowfall was proposed, Firstly, snowfall model and snow surface effect were analyzed in detail, the snow model was applied to image segmentation with strong adaptability, and then mixed the traditional random walk image segmentation algorithm with adaptive snow model characteristics, generated a new algorithm, finally made performance simulation using virtual and real images algorithm, the results showed the image segmentation performance is better than the common NCut and the traditional random walk algorithm for image segmentation, and it had certain research value.**

***Index Terms*—Image Segmentation; Snowfall Model; Random Walk; Gauss Kernel Function**

I. INTRODUCTION

As one of objectives on computer visualization, the image has been researched frequently in recent years, such as meteor gram, remote sensing monitoring image, two-dimension code, etc. all of which contains much and is the vehicle of information. As the vehicle of information, the foreground and background of the image becomes familiar among people in terms of its visual characteristics, in a certain case, there is of a necessity to abstract the small image from the original image without loss to take selective analysis, generally speaking, the small image is irregular, which needs the image segmentation for the abstraction. As the key step of image analysis, the image segmentation plays an important role in the image process.

Image segmentation is not the simple image cropping, during the image segmentation, it should pay a special attention to the edge process, for more edge cropping will cause noise, and less edge cropping will lose the pictorial information. Therefore, the key to be solved in the image segmentation is how to keep the nature of original image as well as causing noise as little as possible. Currently, many scholars has done research on the image segmentation, all of which are algorithm optimization

basically, and there is no common way to make the anti-noise performance and detail-keeping reach the best, for example, the image segmentation realized by the fuzzy clustering, which is made through the C mean value algorithm of fuzzy clustering in the literature [3], but FCM algorithm in literature [4-5], objectively, some certain but not outstanding effects will be med through the fuzzy clustering algorithm in the complex image segmentation. The mean shift is applied in the remote sensing image in the literature [6]; the image segmentation of self-adapting random walk is applied in the literature [7], which has a good effect but still needs improvement.

This paper based on the simulation of the snow image segmentation, new ideas, exploration has the characteristics of image region under the guide of image contour curve evolution and visual characteristics of the image contour extraction method that accords with people. Adaptive gaussian kernel data can achieve adaptive according to image feature information region of the image surface evolution, in view of this, this article introduce the adaptive gaussian kernel data model, the snow to try to simulate simple snow model, namely the adaptive evolution of snow surface model. Similar to simulate snowfall process, based on the model of snow surface implied the ground contour curve of the evolution of the surface, and a topological adaptability. The discussion and the research provide a new research idea for the extraction of image contour.

The focus is the evolution of snow surface effect, realize the snow model. Snow model here is obtained by facet model to improve the traditional method to estimate, namely by two-dimensional surface fitting surface evolution, and introduces the data adaptive Gauss kernel, data adaptive Gauss kernel according to the characteristics of the image area information from the image surface evolutionary adaptation, image processing window of the improved model according to the characteristics of adaptive image itself and change the size, shape and direction, and this is an iterative process.

An image process way based on the snowfall model is put forward in the paper, the focus of which is to imitate

the image as the rolling ground, and then imitate the landmark curve changing similar to the process of snowfall. The snowfall model is shaped as the self-adopting model based on the earth surface, so it can be well applied into the image segmentation, reduces the edge noise effectively and ensures the elegance and smoothness of the image segmentation. The snowfall model is merely an optimizing process for the edge of image segmentation, which also needs the algorithm of random walk to complete the image segmentation. It has proved that the image segmentation made by such a model is more precise.

II. PROPOSED SCHEME

The surface feature is quantized by the gradient in the analysis of snowfall model; in the process of snowfall, the snowflake falls onto the ground, then the surface feature will change inordinately as well as the surface outline, therefore, the surface outline can be as the cut-off rule of image segmentation. That is to say, the snowfall process is imitated in the abstracting process of image segmentation, according to the continuous change of surface curve, the segmentation line is abstracted and the image segmentation is realized.

A. The Surface Effect of Accumulated Snow

The surface effect of accumulated snow will directly affect the vision, and the accumulated snow will change the image information of object itself inordinately, firstly, the snow will cover part of the object surface, which will lower the hardness of edge abstraction and highlight the entire outline of the object if the image of object is abstracted from the whole; secondly, the accumulated snow effect will effectively reduce the setoff relations among the objects, eliminate the unnecessary shadow and wipe out the noise; finally, the snow will highlight the outline of large objectives while obscuring the details.

The surface effect of snowfall model may have a certain relation with the time of snowfalling, so such a relation should be considered when processing the image segmentation; the setting of coefficient of snowfalling not only ensure the time of snowfalling and refrain the noise as far as possible, but also prevent a too-long time of snowfall, and overlay the beneficial information, therefore, there is a need to weigh and consider the balance, and set up a reasonable weight coefficient of snowfalling, and such a coefficient of snowfalling is limited by the times of snowfalling in the process of simulation process.

B. The Basic Snowfall Model

The snowfall effect can change the surface outline, and simulate the self-adopting surface curve; a key needs to be solved for the modeling of snowfall is how to quantize the surface curve by virtue of mathematical way; the Gaussian kernel function will be applied in quantizing the surface change of snowfall.

In the following, the author will improve the primary algorithm of Gaussian kernel function, so as to better satisfy the actual process of image segmentation; the diagram is used to describe the core thought of algorithm to illustrate the answer visually. Figure 1. shows that any image will be shaped after the standard image has been stretched, rotated, and processed by changing of scale; characteristics of any image can be expressed by the parameter of standard image, stretch parameter Λ_i, spin parameter $\mathbf{U}_{\theta_i}$ and parameter of changing of scale γ_i. Therefore, any characteristics of image can be completed by the self-adopting Gaussian kernel.

Seen from figure 2 that the shape of different kernel function is described, in which (a) is the traditional Gaussian kernel, describing the flat side in the image, (b) describes the texture area, in which the image will be abundant after the image becomes small; (c) describes the weak fringe area; (d) describe the strong fringe area; (e) describes the arris edge area. The brief introduction of analysis graphics 3 of classical image is applied to visually illustrate the characteristics of the five areas.
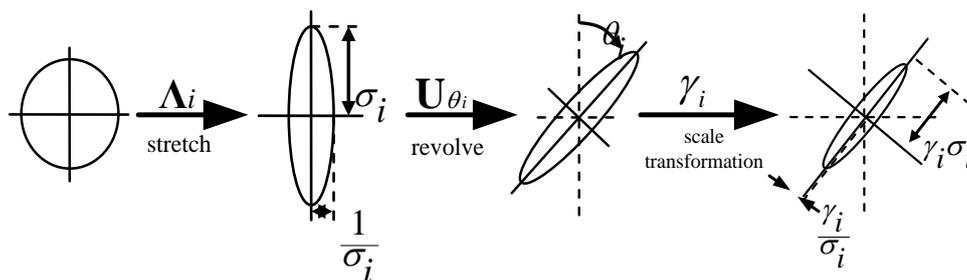
Figure 1.   The improvement of traditional kernel function
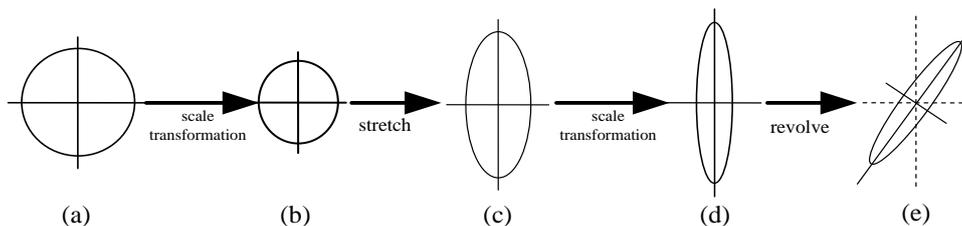
Figure 2.   The evolution kernel of self-adopting curve surface
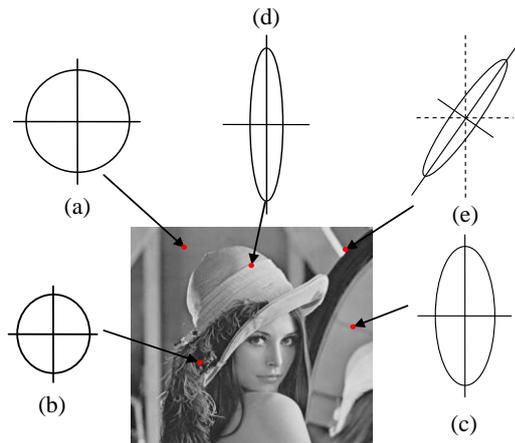
Figure 3. The schematic diagram of curve self-adoption

The self-adoption Gaussian kernel function is put forward in accordance to the conversion process described above:

$$k(y, y_0) = \exp\left(-\frac{(y - y_0)^{\mathrm{T}} \mathbf{C}_0 (y - y_0)}{2\sigma^2}\right) \quad (1)$$

There into σ describes the whole smooth factor, and C_0 describes the covariance matrix based on different partial gray value; eigenvalue decomposition is made for C_0 so as to simplify the calculation:

$$\mathbf{C}_0 = \gamma \mathbf{U}_{\theta_0} B_0 \mathbf{U}_{\theta_0}^{\mathrm{T}}$$
$$\mathbf{U}_{\theta_0} = \begin{bmatrix} \cos(\theta_0) & \sin(\theta_0) \\ -\sin(\theta_0) & \cos(\theta_0) \end{bmatrix} \quad (2)$$
$$B_0 = \begin{bmatrix} \sigma_0 & 0 \\ 0 & \sigma_0^{-1} \end{bmatrix}$$

There into U_{θ_0}, B_0, σ_0 describes the rotation matrix, stretching rectangle and the size of kernel after adjustment, respectively. The gradient matrix can be defined as follows:

$$M_0 = \begin{bmatrix} \vdots & \vdots \\ I_{y_1}(y_j) & I_{y_2}(y_j) \\ \vdots & \vdots \end{bmatrix} = \mathbf{U}_0 N_0 \mathbf{V}_0^{\mathrm{T}}, \ y_j \in w_0 \quad (3)$$

There into $\mathbf{U}_0 N_0 \mathbf{V}_0^{\mathrm{T}}$ is the decomposition of M_0, and N_0 is the diagonal matrix of second order, which is the direction value to be a key; $V_0, V_0 = [v_1, v_2]^T$ is a orthogonal matrix to define the direction, in which the angle value is

$$\theta_0 = \arctan\left(\frac{v_1}{v_2}\right) \quad (4)$$

The rotation parameter σ_0 is defined by the determinative gradient direction:

$$\sigma_0 = \frac{s_1 + \eta'}{s_2 + \eta'}, \ \eta' \geq 0 \quad (5)$$

There into η' is the stretched regulatory factor.

The selective factor ϕ_0 is

$$\phi_0 = \left(\frac{s_1 s_2 + \eta''}{M}\right)^{1/2} \quad (6)$$

There into η'' is the selective factor.

C. Algorithm Design

The random walk is a network generation approach based on the diffusion strategy, the process of which is to install an active label in the entire original network, and the active label's walk process is the process of random walk. The panel point that the active label goes through is the panel point of reconstructed network; the random walk strategy doesn't mean selecting a panel point randomly in the connected panel points, but selecting next moving or withdrawing back to the original panel point based on the network attribution of panel point, that is, springing strategy, which can effectively avoid the partial lost, meanwhile, what needs to be considered is the incomplete ergodic of original network, so as to avoid the uneven sample of the entire network.

Dirichlet problem has similarities with the solution of transition probability of random walk in the case of fixed boundary conditions, hereon, the detailed description of solution process of random walk algorithm will be made by virtue of Dirichlet problem:

The dirichlet integral in the given area Z is

$$D[u] = \frac{1}{2} \int_\Omega |\nabla u|^2 dZ \quad (7)$$

Starting from the non-gauge point, the probability that reaches the gauge point for the active label of random walk is equal to the solution of Dirichlet problem under the condition of boundary T, in which $u(i, j)$ is the harmonic function, and the boundary conditions is described as equation (3):

$$\nabla^2 u = \frac{\partial^2 u}{\partial i^2} + \frac{\partial^2 u}{\partial j^2} = 0 \quad (8)$$

$$u(\mathrm{i,j})\big|_\Gamma = \begin{cases} 1 & (i, j) = s \\ 0 & others \end{cases} \quad (9)$$

The solution process of the entire problem is the solution process of minimum value of $u(i, j)$.

The united Laplacian matrix is defined in the mapped picture described in equation (4):

$$L_{ij} = \begin{cases} d_i & i = j \\ -w_{ij} & v_i \text{ adjoin } v_j \\ 0 & others \end{cases} \quad (10)$$

The value of L_{ij} is codetermined by the panel points v_i and v_j, and d_i is the point of v_i. The incidence matrix between peak points is the definition of $m \times n$ in figure G.

$$A_{e_{ij}v_k} = \begin{cases} +1 & if\ i = k \\ -1 & if\ j = k \\ 0 & others \end{cases} \quad (11)$$

Gained from the equation (11) that the incidence matrix is codetermined by the edge e_{ij} and the panel point v_k; all the e_{ij} in the figure can be an arbitrarily given direction. Generally, A is called united gradient operator and A^T is called united divergence operator.

To define a opposite angles with the size of $m \times m$, and the matrix C is structured, the catercorner value of which is equal to the weight of the edge in the corresponding mapped picture, that is,

$$C_{e_{ij}e_{ks}} = \begin{cases} w(e_{ij}) & i = k, j = s \\ 0 & others \end{cases} \quad (12)$$

Based on the premise of continuity, all the isotropous Laplacian matrix can be decomposed into the product of the joint divergence operator and the joint gradient operator, that is, $L = A^T A$, and such a structure matrix can be understood as the measurement of a vector quantity plus the weight inner product, in such a sense, when $C = I$ through $L = A^T CA$, then $L = A^T A$.

Under the premise of fixed gauge point, the probability value that the non-gauge point reaching the gauge point is solved. The equation (7) can be rewritten as

$$D[y] = \frac{1}{2}(Ay)^T C(Ay)$$
$$= \frac{1}{2}y^T Ly = \frac{1}{2}\sum_{e_{ij} \in E} w_{ij}(y_i - y_j)^2 \quad (13)$$

$$D[y_U] = \frac{1}{2}[y_M^T\ y_U^T]\begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix}\begin{bmatrix} y_M \\ y_U \end{bmatrix}$$
$$= \frac{1}{2}(y_M^T L_M y_M + 2y_U^T B^T y_M + y_U^T L_U y_U) \quad (14)$$

$$L = \begin{bmatrix} L_M & B \\ B_T & L_U \end{bmatrix} \quad (15)$$

As for $D[x_U]$ Take a differential for y_U, then gain the equation (10) that

$$L_U y_U = -B_T y_M \quad (16)$$

Demand y_i^s as the probability value that the non-seed point v_i reaches the seed point of s, and then define such a function to express the assemble of all gauge points:
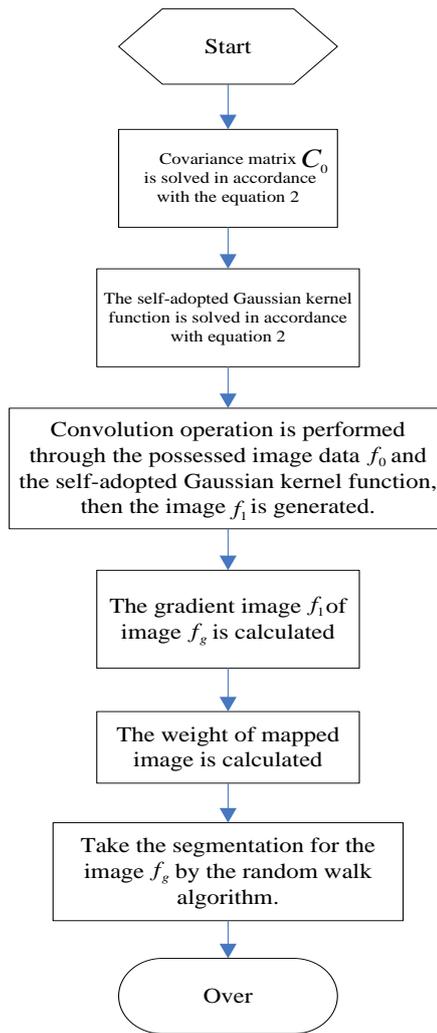
Figure 4.   The flow diagram of image segmentation of random walk based on the snowfall model

$Q(v_j) = s, \forall v_j \in V_M, and\ s \in Z, 0 < s \le K$, then define all the points(belonging to $v_j \in V_M$) to be a matrix $|V_M|_{\times 1}$.

$$m_j^s = \begin{cases} 1 & if\ Q(v_j) = s \\ 0 & if\ Q(v_j) \ne s \end{cases} \quad (17)$$

$$L_U y^s = -B^T m^s \quad (18)$$

$$L_U Y = -B^T N \quad (19)$$

By virtue of $K - 1$ linear equation, the probability value that non-gauge points reaching to K gauge points is solved, and K lines for X is gained by y^s, moreover, the line number of M can be gained by m^s; seen from the circuitous philosophy, the probability sum of each points are 1.

$$\sum_s y_i^s = 1, \quad \forall v_i \in V \quad (20)$$

After the probability with the amount of K that each non-gauge point v_i reaches gauge points has gained,

compare them with their size, and take the $\max_s(x_i^s)$ as the principle to realize the image segmentation.

The Gaussian function is applied into the traditional random walk to determine the weight of the edge composing of two pixels, that is,

$$w_{ij} = \exp(-\beta(g_i - g_j)^2) \qquad (21)$$

The gradient information has been applied when the weight function is selected in the paper, that is,

$$w_{ij} = \exp[-\beta_1(g_i - g_j)^2 - \beta_2((h_i - h_j)^2)] \quad (22)$$

g_i is the gray value of image pixel i after the snowfall model has been processed; h_i is the corresponding gradient value; β_1, β_2 is the free parameter to improve the algorithm.

It can be gained through the analysis above that the new image segmentation algorithm is based on the fusion of the image segmentation algorithm of traditional random walk and the characteristics of self-adopted snowfall model, and the realization process of detailed algorithm is shown as figure 4.

III. SIMULATION RESULTS

In order to review the function of algorithm of improve random walk based on the snowfall model put forward in the paper, the Matlab and the algorithm listed in the section 2.3 is applied to take the simulation of examples for the manual simulation image and the real image, then compare the NCut and random walk in terms of the function.

(A) the original image (B) the gradient image (C) NCut (D) random walk (E) the paper

(F) noise-added image (G) gradient image (H) NCut (I) random walk (J) the paper
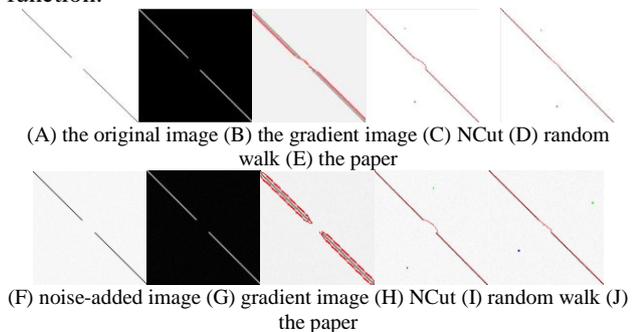
Figure 5. The comparison of segmentation of virtual image (a-e) without added noise; (f-j) adds the Gaussian noise

A. Virtual Image

Firstly, the Matlab is applied to simulate the straight line image with a cut, and the algorithm of random walk of self-adopted snowfall model is to perform the image segmentation, which aims to prove the segmentation of incomplete image through such an algorithm. The simulation image is shown through figure 5(a); (b) is the gradient image that has been filtered after the snowfall model; (c) is the image that has been segmented by the normalization; (d) is the image of traditional random walk; (e) is the segmentation image of random walk of self-adopted snowfall model applied in the paper. Seen from the figure, the make-up function of the algorithm for the incomplete figure applied in the paper is more

powerful; the image cutting-off line cannot be coincided with the straight line; the segmentation in (d) has a little fraction of intrusion in the incomplete part.

(1)

(2)

(3)
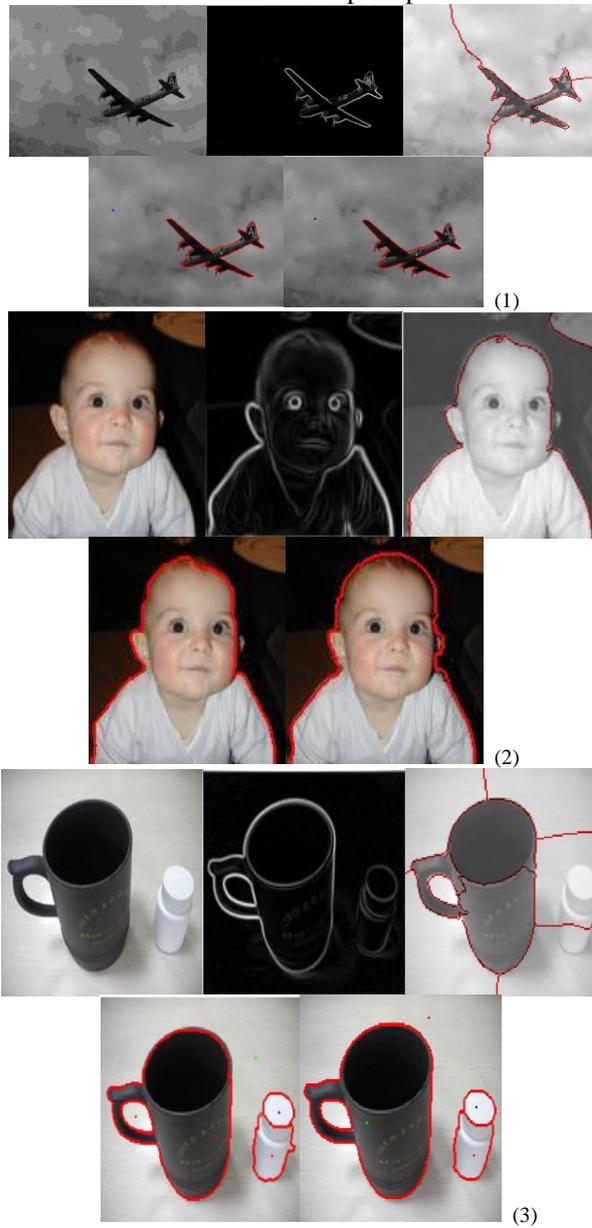
(a) original image (b) gradient image (c) NCut (d) random walk (e) an improve way

Figure 6. The comparison of the segmentation result of real image

Meanwhile, in order to take an analysis and comparison for the anti-noise function of the new algorithm, the Gaussian noise is added into the virtual image (a), in which the mean value and the variance are 0 and 0.01 respectively, shown as figure (f). Correspondingly, (h), (i) and (j) are the segmentation effect after the NCut, traditional random walk and the algorithm in the paper have been used. Seen clearly from the experimental result, the algorithm of random walk based on the self-adopted snowfall model performs better than other two ways in terms of the anti-noise, moreover, the experiment has proved that the stronger the noise sounds, the clearer the superiority of the algorithm used in the paper.

B. Real Image

The plane image, human image and medicine bottle are selected as the experimental objective for the experiment in the paper so as to better illustrate the generality of the algorithm used in the paper, in which the former two images are from the public database, and the third one is shot by the researcher in the daily life. Shown in the figure 6, there are three groups, and each has five images, that is, original image, gradient image, NCut segmentation image, segmentation image of traditional random walk, and image generated by the improved way in the paper.

The images of the first group and the second group are from the segmentation database of Berkeley; within the Matlab simulation, the image of Gaussian noise with mean value 0 and variance 0.0005 has been added into the original image, then the simulation result can be gained; (c) NCut has an ordinary segmentation result; (d) has a certain segmentation result; but the omitted segmentation appears in the first group, and the segmentation on the entire image is perfect in the paper.

The third group is the user-defined image, although the background of image is easy, the shadow between two objects increase the hardness for the two image segmentation. Shown in the figure (c), the segmentation result of NCut is bad, and the segmentation result of the cup on the left is unclear, what is worse, the medicine bottle fails to be identified due to the mistakes for the background; the segmentation result of the right objects is relative bad in the (d) segmentation algorithm, in which the burrs appear. The advantage of segmentation result used in the paper is relatively good.

IV. CONCLUSION

The image segmentation algorithm of random walk of self-adopted snowfall model has been put forward in the chapter, which has proved through experiments that the combination of the traditional random walk and the snowfall model can better improve the image segmentation, especially the edge process of the image segmentation can not only maintain the original characteristics, but also better refrain the noise, therefore, it has a certain application prospect.

The follow-up study mainly aims at optimizing the algorithm function, and further improving the precision of the self-adopted parameter control of the snowfall model, and improving the segmentation integrity of the complex image.

ACKNOWLEDGMENT

*Corresponding author. Email address: WJH@vip.163.com

REFERENCES

[1] Wu, YiquanHao, Yabing Wu, Shihua Zhang, Yufei Xie, Qiankun. Marine spill oil SAR image segmentation based on KFCM and improved CV model, *Chinese Journal of Scientific Instrument,* v 33, n 12, December 2012: pp. 2812-2818.

[2] Khan, Ahmad Ullah, Javid Jaffar, M. Arfan Choi, Tae-Sun. Color image segmentation: a novel spatial fuzzy genetic algorithm. *Signal, Image and Video Processing*, 2012 pp. 1-11.

[3] Liu Fang. The new image segmentation algorithm using adaptive evolutionary programming and fuzzy C-means clustering. Proceedings of SPIE - The International Society for Optical Engineering, v 8056, 2011, *Visual Information Processing.*

[4] Li Xuchao, Liu Haikuan, Wang Fei, Bai Chunyan. The fuzzy clustering way in the image segmentation. Journal of Image and Graphics2012, (04), pp. 100-106.

[5] Yi Yufeng, Gao Liqun, Guo Li. The improved application of the FCM in the interactive image segmentation. Chinese journal of image and graphics, 2012, (03) pp. 90-96.

[6] Zhou Jiaxiang, Zhu Jianjun, Mei Xiaoming, Ma Huiyun. The remote sensing image segmentation of the self-adopted meanshift with multi-dimensional characteristics Information. Journal of wuhan university (science edition), 2012, (4) pp. 419-422440.

[7] Yi Yufeng, Gao Liqun, Cheng Wei, Yu Hongyin. The image segmentation algorithm of the self-adopted random walk. *Journal of northeastern university (natural science edition)*, 2011, (8) pp. 1092-1096.

[8] Wang Chengen Huang Zhangjun. The Kriging response surface method based on the Gaussian function and the updating strategy of trust region. *Journal of computer integrated manufacturing system,* 2011, (4) pp. 740-746.

[9] Dou Feiling, Hu Yanqing, Li Yong, Fan Ying, Di Zengjia. The random walk of the space network. *Journal of physics,* 2012, (17) pp. 571-577.

[10] Wang Fangmei, Fan Hong, Wang Fengni. The application research on the level set in the image segmentation. Computer application research, 2012, (4) pp. 1207-1210.

Yan Yu, born in 1976-02-23, Harbin, Heilongjiang province, China. He is currently a Associate Professor in Harbin Normal University. His research interests include knowledge engineering, information retrieval, intelligent education.

Jianhua Wang was born in 1956. Professor and Ph. D. supervisor in Harbin Normal University. Her main research interests include image processing, intelligent information processing.

Madukasi Charles Nnaemeka is Nigerian, born in 1989. He has got degree from school of international exchange, Harbin Normal University. His main research interests include image processing and artificial intelligence.

Human Face Recognition based on Improved PCA Algorithm

Xu Yue

College of art and design, LanZhou JiaoTong University, Lanzhou, China
Email: xuyue_19630408@163.com

Linhao Li

AT&T Labs, 200 South Laurel Ave, #D4-3C05, NJ, USA
Email: lli@research.att.com

Abstract—**This paper aims to effectively recognize human faces from images, which is an important problem in the multimedia information process. After analyzing the related research works, the framework of the face recognition system is illustrated as first, which contains the training process and the testing process. Particularly, the improved PCA algorithm is use in the feature extraction module. The main innovations of this paper lie in that, in the improved PCA, we utilize a radial basis function to construct a kernel matrix by computing the distance of two different vectors, which are calculated by the parameter of 2-norm exponential. Afterwards, human faces can be recognized by computing the distance of test image and the training images by the nearest neighbor classifier, of which the cosine distance is utilized. Finally, experiments are conducted to make performance evaluation. Compared with the existing face recognition methods, the proposed scheme is more effective in recognizing human faces with high efficiency.**

Index Terms—**PCA, Face Recognition, Feature Extracting, Kernel Matrix, Cosine Distance**

I. INTRODUCTION

As there are many terrorist attacks in recent years, various government agencies are now more motivated to improve security data systems based on body or behavioural characteristics. On the whole, biometric systems process raw data to extract a template which is easier to process and store, but carries most of the information needed. This technology is very attractive, the reason lies in that this technology can be combined into other applications related to security management or access control. Unlike other biometrics recognition, face recognition belongs to a good compromise between reliability and social acceptance and balances security and privacy well. It can be seen that any identification system utilizing face recognition technology poses several threats to civil rights, because it impinges on the privacy of innocent people when false positives are investigated [1].

As is illustrated in Wikipedia, Face recognition refers to an individual's understanding and interpretation of the face, particularly the human face, especially in relation to

the associated information processing in the brain. Face recognition systems fall into two categories: verification and identification. Face verification is a 1:1 match that compares a face image against a template face images, whose identity is being claimed. On the contrary, face identification is belonged to a 1: N problem which compares a query face image against all image templates in a face database to determine the identity of the query face. The test individual may or may not be in the system database. The query face image is compared against all the face images in the database, computing a score for each one. All these scores are numerically ranked so that the highest score is first, and if a similarity score is higher than a given threshold, an alarm is raised [1-3].

In last recent years, there are many developments occurred in the research field of face recognition, with many systems capable of obtaining recognition rates greater than 90%. However real-world scenarios remain a challenge, because face acquisition process can undergo to a wide range of variations.

Face recognition is an important step of the capability of human perception system and is a routine task for humans, while building a similar computer system is still far from satisfactory. The earliest work on face recognition began in 1950s in the research field of psychology, on the other hand, in 1960s this research are turning to engineering literature. The problem of automatic face recognition contains three main steps, which are: 1) detection and rough normalization of faces, 2) feature extraction and accurate normalization of faces, and 3) identification and/or verification. Usually, different sub-tasks are not totally separated. For example, the facial features (such as ears, eyes, nose, mouth et al.) utilized for face recognition are often used in face detection. Face detection and feature extraction can be achieved simultaneously. Depending on the nature of the application, for example, the sizes of the training and testing databases, clutter and variability of the background, noise, occlusion, and speed requirements, some of the sub-tasks can be very challenging [4-7].

The main innovations of this paper lie in the following aspects:

(1) We proposed a novel improved PCA algorithm which could utilize a radial basis function to construct a kernel matrix by computing the distance of two different vectors, which are computed by the parameter of 2-norm exponential.

(2) We use the sliding median filter in dealing with signals which could contaminate with outliers, and the relationship one dimension sliding median filter between input and output in this window of the sliding median filter is also considered.

(3) The Kronecker production, the wavelet decomposition technology and the sliding median filter are also utilized in the proposed scheme.

(4) We implement the improved PCA algorithm and utilize it in calculating the column data matrix and the corresponding eigenvectors.

The rest of the paper is organized as the following sections. Section 2 introduces the related works. Section 3 illustrates the proposed scheme for recognizing human faces from images based on an improved PCA algorithm. In section 4, a series of experiments are designed and conducted to make performance evaluation. Finally, we conclude the whole paper in section 5.

II. RELATED WORKS

As face recognition is an important problem in image processing, in recent years, there are many related research works in this field. In the following section, we will introduce some typical works.

Kan et al. propose Adaptive Discriminant Analysis in which the within-class scatter matrix of each enrolled subject is inferred using his/her single sample, by leveraging a generic set with multiple samples per person. The proposed method is motivated from the assumption that subjects who look alike to each other generally share similar within-class variations. In ADA, a limited number of neighbors for each single sample are first determined from the generic set by using kNN regression or Lasso regression. Then, the within-class scatter matrix of this single sample is inferred as the weighted average of the within-class scatter matrices of these neighbors based on the arithmetic mean or Riemannian mean [8].

Tang et al. developed a novel 3D face recognition algorithm using Local Binary Pattern under expression varieties, which is an extension of the LBP operator widely used in ordinary facial analysis. First, to depict the human face more accurately and reduce the effect of facial local distortion for face recognition, a special feature-based 3D face division scheme is proposed. Then, the LBP representation framework for 3D faces is described, and the facial depth and normal information are extracted and encoded by LBP, to reduce the expression effect. For each face region, the statistical histogram is utilized to summarize the facial details, and accordingly three matching strategies are presented to address the recognition task [9].

Feng et al. proposed to learn jointly the projection matrix for dimensionality reduction and the discriminative dictionary for face representation. The joint learning makes the learned projection and dictionary

better fit with each other so that a more effective face classification can be obtained [10].

Shafey et al. present a scalable and exact solution for probabilistic linear discriminant analysis (PLDA). Moreover, the authors present a scalable derivation which is theoretically equivalent to the previous nonscalable solution and thus obviates the need for a variational approximation [11].

Yang et al. presented a Gabor feature based robust representation and classification scheme is proposed for robust FR. The use of Gabor features not only increases the discrimination power of face representation, but also allows us to compute a compact Gabor occlusion dictionary which has much less atoms than the identity occlusion dictionary. Furthermore, the authors testify that with Gabor feature transformation, $l(2)$-norm could take the role of $l(1)$-norm to regularize the coding coefficients, which reduces significantly the computational cost in coding occluded face images [12].

Xiao et al. proposed a novel method for face recognition by integrating texture information with shape information, called biview face recognition algorithm. The texture models are constructed by using subspace learning methods and shape topologies are formed by building graphs for face images. The proposed biview face recognition method is compared with recognition algorithms merely based on texture or shape information [13].

Pedraza et al. developed a face recognition system based on soft computing techniques, which complies with privacy-by-design rules and defines a set of principles that are context-aware applications and should contain to conform to European and US law. Particularly, this paper deals with the necessity to consider legal issues concerning privacy or human rights in the development of biometric identification in ambient intelligence systems [14].

Dornaika et al. introduced a novel discriminant technique called "exponential LDE". The proposed scheme can be seen as an extension of LDE framework in two directions. First, the proposed framework overcomes the SSS problem without discarding the discriminant information that was contained in the null space of the locality preserving scatter matrices associated with LDE. Second, the proposed ELDE is equivalent to transforming original data into a new space by distance diffusion mapping, and then, LDE is applied in such a new space [15].

Yang et al. proposed a novel robust kernel representation model with statistical local features (SLF) for robust face recognition. Initially, multipartition max pooling is used to enhance the invariance of SLF to image registration error. Then, a kernel-based representation model is proposed to fully exploit the discrimination information embedded in the SLF, and robust regression is adopted to effectively handle the occlusion in face images [16].

In this paper, Principal Component Analysis (PCA) is used in recognizing human faces, which is a powerful

computing tool in intelligent computing. The application of PCA in intelligent computing is shown as follows.

Wan et al. proposed a novel image fusion scheme for combining two or multiple images with different focus points to generate an all-in-focus image. The authors formulate the problem of fusing multifocus images as choosing the most significant features from a sparse matrix obtained by a newly developed robust principal component analysis (RPCA) decomposition method to form a composite feature space. The local sparse features that represent salient information of the input images (i.e. sharp regions) are integrated to construct the resulting fused image [17].

Liang et al. propose Lp-norm generalized principal component analysis by maximizing a class of convex objective functions. The successive linearization technique is used to solve the proposed optimization model. It is interesting to note that the closed-form solution of the sub-problem in the algorithm can be achieved at each iteration. Meanwhile, they theoretically prove the convergence of the proposed method under proper conditions [18].

Hajati et al. studied the influence of spectral noise on a new method for three-dimensional X-ray photoelectron spectroscopy imaging, which is based on analysis of the XPS peak shape. Particularly, they study in more detail the influence of noise reduction by principal component analysis on 3D XPS images of carbon contamination of a patterned oxidized silicon sample and on 3D XPS images of Ag covered by a nano-scale patterned octadiene layer [19].

Castano et al. proposed a deterministic algorithm to initiate any hidden node with an additive activation function to be trained with ELM. The proposed algorithm uses the information retrieved from principal components analysis to fit the hidden nodes. This approach considerably decreases computational cost compared to later ELM improvements and overcomes their performance [20].

Based on the above analysis, in this paper, we concentrate on how to recognize human faces from images based on an improved PCA algorithm.

III. THE PROPOSED SCHEME

A. Framework of the Face Recognition System

In this section, we will show a framework of the face recognition system based on modified PCA algorithm (shown in Fig. 1). This proposed framework is made up two module, which are training process and testing process. In the training process, we utilize Adaboost algorithm to obtain 2-D gray face images. Afterwards, the face images are enhanced by histogram equalization. In this testing process, after the image enhancement, face detecting and feature extracting module, face recognition results are obtained by the classifier which is trained by training process.

B. The Improved PCA

However, in some complex cases in industrial chemical and environmental processes, which especially

have nonlinear characteristics, PCA exhibits bad behavior because of its linearity assumption [22]. Therefore, in order to enhance the accuracy of face recognition, in this paper, we propose an improved PCA algorithm.

We utilize a radial basis function to construct a kernel matrix by computing the following equation as follows.

$$K_j^i = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{c}\right) \qquad (1)$$

where the distance of two different vectors are computed by the parameter of 2-norm exponential.

Supposing $X \in R^{M \times N}$ refer to a sample matrix, given a new test sample $S \in R^M$, the kernel vector is calculated as follows.

$$v = \lambda(S)^T \cdot \lambda(X) \in R^N \qquad (2)$$

Supposing $C \in R^{N \times N}$ is the coefficient matrix, and it can be obtained by the following equation.

$$C = \frac{t}{\left(\frac{t^T \cdot K \cdot t}{N}\right)^{\frac{1}{2}}} \in R^N \qquad (3)$$

where C is obtained in the period of modeling, and C can be utilized to calculate the new test sample's score.

$$\hat{t} = v \cdot C \in R^N \qquad (4)$$

where $\hat{t}$ denotes the new sample of t. Particularly, t_i represents the ith column of the block score matrix. In order to calculate the SPE statistic, it is of great importance to compute the value of $\lambda(S)$. This process is implemented as follows.

$$\lambda(S) = \hat{t} \cdot U^T = \hat{t} \cdot C^T \cdot \lambda(S)^T \qquad (5)$$

where $U = \lambda(S) \cdot C$. Next, the SPE statistic is obtained by the following equation.

$$\begin{aligned} SPE &= \left\|\lambda(S)^T - \lambda(S)\right\|^2 \\ &= \lambda(S)^T \cdot \lambda(S) - 2 \cdot \lambda(S) \cdot \lambda(S) + \lambda(S) \cdot \lambda(S)^T \\ &= k(S,S) - 2 \cdot \hat{t} \cdot P^T \cdot \lambda(S) + \hat{t} \cdot P^T \cdot P \cdot \hat{t}^T \\ &= 1 - 2 \cdot \hat{t} \cdot C^T \cdot v^T + \hat{t} \cdot C^T \cdot K \cdot C \cdot \hat{t}^T \end{aligned} \qquad (6)$$

Afterwards, the sliding median filter is utilized to deal with signals which could contaminate with outliers, and the relationship one dimension sliding median filter between input and output in this window of the sliding median filter is given as follows.

$$output(k) = med(intput(k-1), \cdots, intput(k), intput(k+1)) \qquad (7)$$

Based on the above analysis, the Kronecker production, the wavelet decomposition technology and the sliding median filter are also contained.
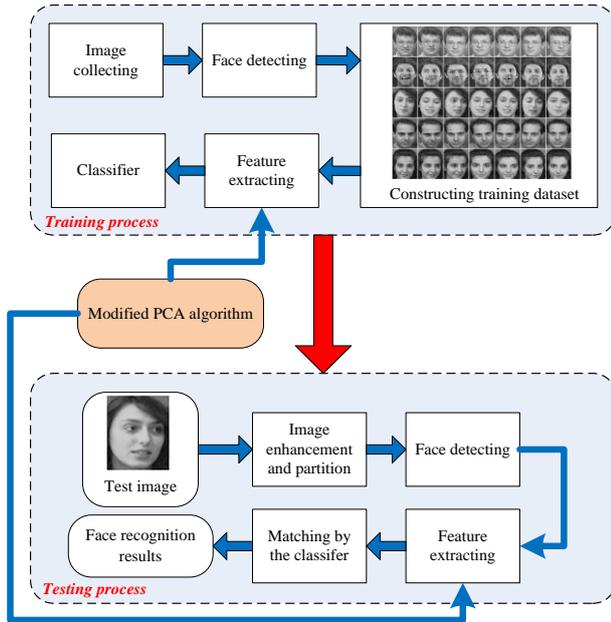
Figure 1.   Framework of the face recognition system based on improved PCA algorithm

C. Face Recognition Algorithm based on the Improved PCA

In the face recognition process, we assume there are N images with human faces which are belonged to M people in the given training image dataset. The process of the proposed algorithm is shown in Fig. 2. As a pre-processing unit, each images are equally divided into several blocks, the size of which is fixed to Z. Particularly, there no two blocks which are overlapped with each other. By this way, each image can be represented as a column data matrix as follows.

$$Y_i = \{y_{i1}, y_{i2}, \cdots, y_{iN}\}, i \in [1, Z] \quad (8)$$

Afterwards, we delete the mean value of each column of the proposed column data matrix. Hence, the modified vertically centered matrix is shown in the following equation.

$$Y_i = \{y_{i1}, y_{i2}, \cdots, y_{iN}\}, i \in [1, Z] \quad (9)$$

In this representation, each column of the matrix is defined as a zero-mean sub-pattern face image.

Based on the above analysis, the modified PCA algorithm can be implemented on the column data matrix and then corresponding eigenvectors can be obtained. Afterwards, Z eigenvector's sets can be calculated, and each element of which is represented as a eigenvector matrix as follows.

$$Q_i = \{q_{i1}, q_{i2}, \cdots, q_{il}\}, i \in [1, Z] \quad (10)$$

where $q_{i1}, q_{i2}, \cdots, q_{il}$ represent the orthonormal matrix, which is related to the first l eigenvalues with largest positive values. The related sub-feature weights using Q_i are calculated as follows.

$$W_i = Q_i^T Y_i = \{w_{i1}, w_{i2}, \cdots, w_{iN}\}, i \in [1, Z] \quad (11)$$

Next, we integrate all the Z sub-features into one global feature as follows.

$$I(G_k) = \left(\alpha_1 \cdot g_{1k}^T, \alpha_2 \cdot g_{2k}^T, \cdots, \alpha_Z \cdot g_{Zk}^T \right)^T, k \in [1, N] \quad (12)$$

where $\sum_{i=1}^{Z} \alpha_i = 1$, and these parameter are used as weight to dynamically modify the influence of each sub-feature, and $I(G_k)$ represents the global feature vector of the kth face image. For a given test image, we also use the same steps to represent them into Z sub-pattern vectors.

Based on the above analysis, human faces can be recognized by computing the distance of test image and the training images by the nearest neighbor classifier. Particularly, the cosine distance is used to estimate image distance as follows.

$$D(I_{test}, I) = \frac{I_{test}(G_k) \cdot I(G_k)}{\|I_{test}(G_k)\| \cdot \|I(G_k)\|} \quad (13)$$
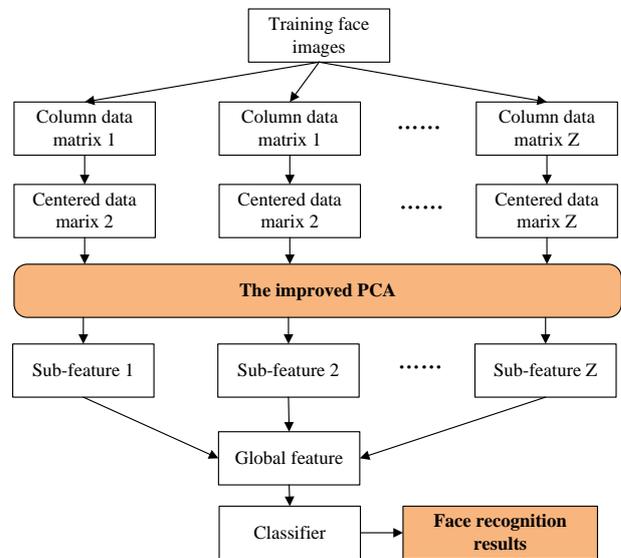
Figure 2.   Process of the proposed face recognition algorithm

IV.   EXPERIMENTAL RESULTS

In this section, we will make performance evaluation for the proposed scheme and compared with other related works. Particulary, we use two face image databases in this experiment, that is, the FERET database and the ORL database are used to compare the proposed approach with the following algorithms: Standard PCA(SPCA), LDA [25], 2DPCA [26], BDPCA [27], 2DLPP [28]. Next, we will conduct the experiment on the ORL and FERET dataset respectively.

A. Experiments Conducted on the ORL Database

ORL database refers to a well-known face database which is made up of face images from 40 individuals, each providing 10 different images. The pose, expression and facial details variations are also extracted. The ORL dataset enables us to evaluate the generalization ability of

proposed methods under circumstances containing varying pose, expression, and face scale, concurrently. All face images in ORL database are normalized to a resolution of 56×46 pixels. Moreover, number of blocks per image NB is set to four. The effect of number of blocks on recognition rate is researched in the following aspects. [21] Examples of the face images in ORL database are shown in Fig. 3.

Figure 3.   Examples of the face images in ORL database.

In this sub-section, the first images of each class are used for training and the remainder ones are utilized for testing. Particularly, the number of training images is chosen as 5, 6, and 8 respectively.
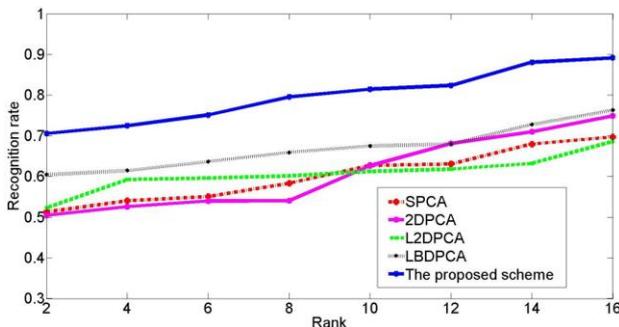
Figure 4.   Recognition rate under the number of training images is equal to 5
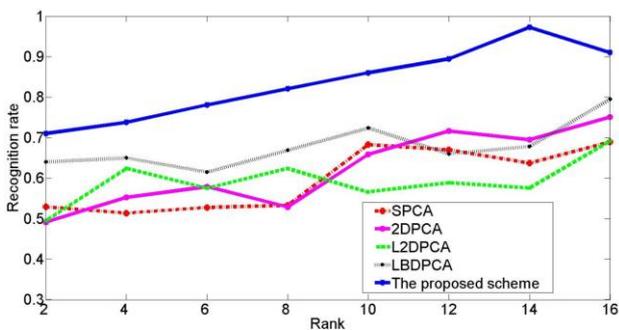
Figure 5.   Recognition rate under the number of training images is equal to 6

From Fig. 4-Fig. 6, we can see that under the performance evaluation metric "Recognition rate", the

proposed scheme perform better than other schemes and the performance of the proposed scheme is better when the number of training images is larger.
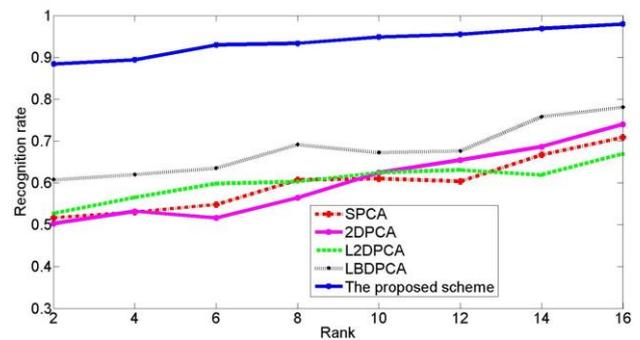
Figure 6.   Recognition rate under the number of training images is equal to 7

In order to show the time cost of the above several method, we save the computing time for each method under the same computer hardware settings, and the experimental results are shown in Table.1, where the parameter N denotes the number of training images. Particularly, the computers in this experiment are installed with Intel Corel i7 CPU, the main frequency of which is 2.9GHz. The memory we used is the 8GB DDR memory with 1600MHz, and the hard disk we utilized is 500GB SSD disk. The graphics chip is the NVIDIA Optimus NVS5400M. Furthermore, the operating system is Windows7 Home Premium (64bit).

TABLE I.   TRAINING TIME AND TESTING TIME CONDUCTED THE ORL DATASET (SECOND).

Scheme	$N = 5$	$N = 6$	$N = 7$
SPCA	4.56	5.57	6.36
	0.90	*0.92*	*0.78*
2DPCA	1.47	1.51	1.53
	1.55	*1.50*	*1.22*
BDPCA	1.61	1.55	1.79
	0.97	*0.90*	*0.87*
2DLPP	11.08	13.76	14.33
	0.95	*1.02*	*0.79*
The proposed scheme	17.57	24.02	31.21
	0.90	*0.82*	*0.70*

In Table.1 the numbers which are marked by the underline refer to the training time, on the other hand, the numbers which are represented as Italic denote to the testing time. For Table.1, we can see that although the training time of the proposed scheme is larger than other schemes, the testing time of our scheme is almost the same to other methods. Considering the performance of the proposed is better than others, the proposed scheme is worth promoting.

B. Experiments Conducted on the FERET Database

The FERET face image database is obtained by the FERET program, which was sponsored by the US Department of Defense through the DARPA Program. This dataset has been a standard database for testing and evaluating the traditional face recognition algorithms. We use a subset of the FERET database, which contains 1400

images of 200 individuals (each individual has 7 images). It involves variations in facial expression, illumination and pose. In this experiment, the facial portion of each original image was automatically cropped based on the location of eyes and the cropped image was resized to 40×40 pixels [23] [24]. The recognition rate under FERET dataset for different scheme is shown in Fig. 6.
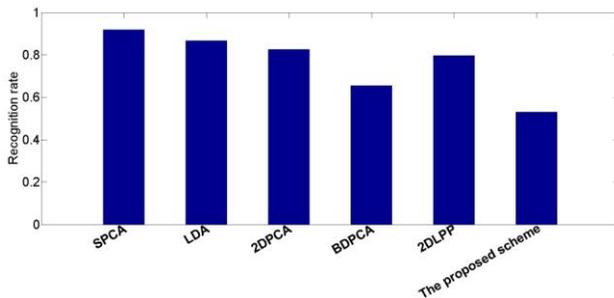
Figure 7. Recognition rate under FERET dataset for different schemes

To give more detailed experimental results, the first several images of each class are chosen as the training images, and then we test the performance under different number of training images. The experimental results are given in Fig. 7, of which the number of training images is 3, 4 and 5 respectively.
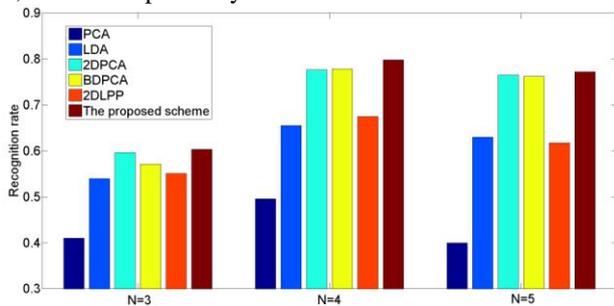
Figure 8. Recognition rate under FERET dataset for different number of training images

C. Experimental Results Analyzing

Based on the above experimental results, in can be seen that the proposed scheme is effective both in the system performance and in the system efficiency. The reasons lie in the following aspects:

(1) The proposed utilize the distribution information of the training images and testing images by the sample similarity weights. However, the BDPCA method does not consider this information.

(2) The 2DPCA method requires more coefficients for image representation than PCA. Although, as a feasible alternative to deal with this problem is to utilize PCA after 2DPCA for further dimensional reduction, the dimensions of 2DPCA can not be reduced directly.

(3) Particularly, BDPCA is directly running on image matrix, however, the standard PCA is required to map an image matrix to a 1-dimension vector at first. On the other hand, BDPCA can solve the problem of classical PCA's over-fitting. Moreover, the ability of feature dimension of BDPCA is much less than 2DPCA.

(4) The main advantages of the 2DLPP algorithm based on LPP lies in that the dimensionality of the bases

derived by the former for image representation is much smaller than the latter. This case means a more accurate approximation of the original images and could also avoid the disadvantages of the existing PCA combined with the LPP algorithm.

(5) The proposed method utilize a radial basis function to construct a kernel matrix by computing the distance of two different vectors, which are computed by the parameter of 2-norm exponential.

(6) In the proposed scheme, the sliding median filter is utilized to deal with signals which could contaminate with outliers, and the relationship one dimension sliding median filter between input and output in this window of the sliding median filter is also considered.

(7) The Kronecker production, the wavelet decomposition technology and the sliding median filter are utilized in this paper.

(8) We use the improved PCA algorithm in computing the column data matrix and the corresponding eigenvectors.

V. CONCLUSIONS

In this paper, we concentrate to design a novel human face recognition, of which an improved PCA is utilized. Particularly, the proposed framework of the face recognition system is made up of the training process and the testing process. Furthermore, the improved PCA algorithm is use in the feature extracting. We utilize a radial basis function to construct a kernel matrix by computing the distance of two different vectors, which are calculated by the parameter of 2-norm exponential in the improved PCA. At last, human faces can be recognized by estimating the distance between the given test image and the training images by the nearest neighbor classifier with the cosine distance.

REFERENCES

[1] Abate Andrea F., Nappi Michele, Riccio Daniel, "2D and 3D face recognition: A survey, Pattern Recognition Letters", 2007, 28(14) pp. 1885-1906.

[2] Wright John, Yang Allen Y., Ganesh Arvind, "Robust Face Recognition via Sparse Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2009, 31(2) pp. 210-227.

[3] Kakadiaris Ioannis A., Passalis Georgios, Toderici George, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(4) pp. 640-649.

[4] Zhao W, Chellappa R, Phillips PJ, "Face recognition: A literature survey, ACM Computing Surveys", 2003, 35(4) pp. 399-459.

[5] Cevikalp H, Neamtu M, Wilkes M, "Discriminative common vectors for face recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(1) pp. 4-13.

[6] Fan Shaosheng, Wang Hainan, "Multi-direction fuzzy morphology algorithm for image edge detection," *Journal of Networks,* 2011, 6(6) pp. 895-898.

[7] Wu Di, Cao Jie, Wang Jinhua, Li Wei, "Multi-feature fusion face recognition based on kernel discriminate local preserve projection algorithm under smart environment",

Journal of Computers (Finland), 2012, 7(10) pp. 2479-2487.

[8] Kan Meina, Shan Shiguang, Su Yu, "Adaptive discriminant learning for face recognition", Pattern Recognition, 2013, 46(9) pp. 2497-2509.

[9] Tang Hengliang, Yin Baocai, Sun Yanfeng, "3D face recognition using local binary patterns", *Signal Processing*, 2013, 93(8) pp. 2190-2198.

[10] Feng Zhizhao, Yang Meng, Zhang Lei, "Joint discriminative dimensionality reduction and dictionary learning for face recognition, Pattern Recognition, 2013, 46(8) pp. 2134-2143.

[11] El Shafey Laurent, McCool Chris, Wallace Roy, "A Scalable Formulation of Probabilistic Linear Discriminant Analysis: Applied to Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2013, 35(7) pp. 1788-1794.

[12] Yang Meng, Zhang Lei, Shiu Simon C. K., "Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary", *Pattern Recognition*, 2013, 46(7) pp. 1865-1878

[13] Xiao Bing, Gao Xinbo, Tao Dacheng, "Biview face recognition in the shape-texture domain, Pattern Recognition, 2013, 46(7) pp. 1906-1919.

[14] Pedraza J., Patricio Miguel A., de Asis A., "Privacy-by-design rules in face recognition system", *Neurocomputing,* 2013, 109 pp. 49-55.

[15] Dornaika Fadi, Bosaghzadeh Alireza, "Exponential Local Discriminant Embedding and Its Application to Face Recognition", *IEEE Transactions on Cybernetics,* 2013, 43(3) pp. 921-934.

[16] Yang Meng, Zhang Lei, Shiu Simon Chi-Keung, "Robust Kernel Representation With Statistical Local Features for Face Recognition", *IEEE Transactions on Neural Networks and Learning Systems*, 2013, 24(6) pp. 900-912.

[17] Wan Tao, Zhu Chenchen, Qin Zengchang, "Multifocus image fusion based on robust principal component analysis", *Pattern Recognition Letters*, 2013, 34(9) pp. 1001-1008.

[18] Liang Zhizheng, Xia Shixiong, Zhou Yong, "Feature extraction based on Lp-norm generalized principal component analysis", *Pattern Recognition Letters*, 2013, 34(9) pp. 1037-1045.

[19] Hajati Shaaker, Walton John, Tougaard Sven, "Three-Dimensional X-Ray Photoelectron Tomography on the Nanoscale: Limits of Data Processing by Principal Component Analysis", *Microscopy and Microanalysis,* 2013, 19(3) pp. 751-760.

[20] Castano A., Fernandez-Navarro F., Hervas-Martinez C., "PCA-ELM: A Robust and Pruned Extreme Learning Machine Approach Based on Principal Component Analysis", *Neural Processing Letters*, 2013, 37(3) pp. 377-392.

[21] Armin Eftekhari, Mohamad Forouzanfar, Hamid Abrishami Moghaddam, Javad Alirezaie, "Block-wise 2D kernel PCA/LDA for face recognition", *Information Processing Letters*, 2010, 110 pp. 761-766.

[22] Lee, Jong-Min and Yoo, ChangKyoo and Lee, In-Beum, "Fault detection of batch processes using multiway kernel principal component analysis", *Computers & chemical engineering*, 2004, 28(9) pp. 1837-1847.

[23] J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss., "The FERET evaluation methodology for face recognition algorithms", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2000, 22(10) pp. 1090-1104.
P. J. Phillips, "The Facial Recognition Technology (FERET) Database", http://www.itl.nist.gov/iad/humanid/feret/feret_master.html, 2004.

[24] Belhumeur Peter N., Hespanha Joao P., Kriegman David J., "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1997, 19(7) pp. 711-720.

[25] Yang Jian, Zhang David, Frangi Alejandro F, Yang Jing-yu, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(1) pp. 131-137.

[26] Zuo Wangmeng, Zhang David, Wang Kuanquan, "Bidirectional PCA with assembled matrix distance metric for image recognition", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2006, 36(4) pp. 863-872.

[27] Hu Dewen, Feng Guiyu, Zhou Zongtan, "Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition", *Pattern recognition*, 2007, 40(1) pp. 339-342.

Xu Yue, He was born in Gansu province, China in 1963 . He graduated from Northwest Normal University in 1982 and is engaged in the design and application of the theory of Arts Education in Colleges and universities, published over 20 academic papers.3 monographs. Won the domestic "literary classic Achievement Award", "gardener Award", "Design Award" and many other awards. Now he is the professor and master's tutor of art and design academy in LanZhou JiaoTong University .

Linhao Li was born in China, May 23, 1978. He received the M.S. in Computer Science from Stevens Institute of Technology (Hoboken, NJ, USA) in 2012. His research interests are in the area of software engineering, wireless network and services orientated computing.

His current job is a Senior Member of Technical Staff at AT&T Labs Research (Middletown, NJ). He has more than 10 years professional experience adept at large scale data analysis and computational complexity.

Mr. Li is a member of IEEE. His team received AT&T Service Excellence Award (2010) and AT&T Labs President Excellence Award (2011). He won 2nd place of "Shanghai Government Software Design Contest" (2004).

Image Hierarchical Representations Models based on Latent Dirichlet Allocation

Fushun WANG and Yan LI

College of Information Science and Technology, Agricultural University of Hebei, Baoding, China

Xiaohua SUN

Department of Digital Media, Hebei Software Institute, Baoding, China

Zhenjiang CAI[1,2]

1. College of Mechanical and Electrical Engineering, Agricultural University of Hebei, Baoding, China;
2. Department of Electrical and Computer Engineering, United States University of Virginia, Virginia, USA)
Corresponding Author: Zhenjiang CAI
Email: fswang_map@163.com

Abstract—**Existing image layer representations methods are very feed-forward, and then not able to deal with small ambiguities. A probabilistic model is proposed, and it learns and deduces each layer in that hierarchy together. Therefore, we consider a recursive probabilistic decomposition process, and derive a new yielded method based on recursive Latent Dirichlet Allocation. We show 2 significant properties of the novel probabilistic method: 1) pulsing another hierarchical to represent the enhanced results on that smooth method; 2) an entire Bayesian method beats a feed-forward running of the novel method. The method can be evaluated on a criterion recognition dataset. It takes the probability of recursive decomposition process into account, and obtains multilayer structure pyramid LDA derived model through the derivation. Experiments demonstrate that the novel technique beats existing hierarchical approaches, and present better performance.**

Index Terms—**Image Hierarchical Representation; Feed-forward; Probabilistic Model; Latent Dirichlet Allocation (LDA)**

I.   INTRODUCTION

Image layer visual representation has been currently concerned in computer vision field. It is not well able to deal with local ambiguities for feed-forward image multilayer visual representation methods. Representing object recognition layer is significant because of the enhanced different, plausibility biologics, and benefited computation. Therefore, Image layer visual representation has been used into many areas such as image retrieval, face recondition and artificial intelligence successfully.

Some articles present a multi-layer structure pyramid LDA model, namely learning and inference hierarchy of all hierarchical probabilistic models to effectively expressing stratified representation of the image. Note: The multi-layer structure pyramid model differs from traditional hierarchical LDA model. Level LDA model is the topic layer formed on the same vocabulary; by contrast, multi-layer structure pyramid LDA model is fixed by recursively formed words, it promotes a single variable of LDA model potential topics to change into Bayesian network variables. Here, we work on level LDA model with meticulous elaboration. Hierarchical LDA model is an extension of the LDA model, the theme depends on the hierarchy. Assuming a given layer L tree, each node is assigned a topic, the text will produce it in the following ways: Choose a pate in the originality of the tree to a leaf node; obtained vector from the L-dimensional Dirichlet in proportion theme; along this route, words resulting from the mixing ratio of the document theme. Finally, the use of CRP (Chinese restaurant process) is to slack assumptions of hybrid tree structure. CRP is an integer division in the distribution, and can be imagined as a process in which the customers of amount M sitting at a Chinese restaurant with limit table number. Main process: the first customer sitting in a seat, then the *m-th* queuing customer seated in the following probability distribution.

Currently, local feature-based description operator, for example SIFT [1], is the most successful and extensive methods in computer vision [2]. It's succeeded in applying to scene and object identification. The main concepts of those methods are: within numerous small areas of some image, coding on the gradient orientation both with compactness and distinction. Based on gained visual words, these local features (clustering center in description space) are able to acquire information. As for this kind of quantization, various scholars put forward different global image representation methods, such as bag-of-words model and space Pyramid mode [3]. And they redefined the measure to replace hard clustering by soft clustering [4]. In other words, it uses individual characteristics of the operator to describe several hybrid features. This method promotes the robustness as well as exists biological system [5]. It's indicated according to the theoretical analysis and biological experimental result

that this decomposition has better performance in data sets of object identification and some new data set [6, 7].

According to the recent research result, multilayer visual presentation could advance the performance of object identification by lifting the metabolic robustness [8], which is exactly corresponding to current stratified theoretical framework and theoretical research of mammal visual cortex [9].These theories extrude the importance of feedback: both improving skills of classification study and explicating local information of reasoning process.

However, most of the existing hierarchical representation model [10-12] process information via feed forward mode, which belongs to strict feed-forward implementation model. It means that every layer's input is the output in former layer which is non-robust for local fuzziness in visual input. In other words, this method cannot solve the problem of local justification. In order to specify local information, we need more contexts expressed by local images. Therefore, it's a necessary to get a recursive Bayesian probabilistic model in allusion to complex visual features.

Based on the above analysis, we propose a probability model for all stratified hierarchy of learning and inference, which is based on being successfully applied to the visual word modeling [13] and the potential Dirichlet distribution (Latent Dirichlet Allocation, LDA) [14] of object detection tasks [7]. It takes the probability of recursive decomposition process into account, and obtains multilayer structure pyramid LDA derived model through the derivation. The model has three important contributions:

We present a multi-layer structure pyramid LDA model, namely learning and inference hierarchy of all hierarchical probabilistic models to effectively expressing stratified representation of the image. To increase the presentation layer can improve the performance of the model plane;

Bayesian probabilistic framework provides a good way to integrate top-down information into the model. Bayesian probabilistic method is superior to feed forward implementation form. An entire Bayesian method beats a feed-forward running of the novel method. The method can be evaluated on a standard recognition dataset.

We choose a pate in the originality of the tree to a leaf node; obtained vector from the L-dimensional Dirichlet in proportion theme; along this route, words resulting from the mixing ratio of the document theme.

In the test results of standard identification data sets show that with the existing hierarchical representation methods, the proposed model showed better performance. In addition, the proposed method has successfully applied into face classification and handwriting recognition.

II. LDA MODEL OF MULTI-LAYER STRUCTURE PYRAMID

This paper presents a multi-layer structure pyramid LDA model, namely learning and inference hierarchy of all hierarchical probabilistic models to effectively expressing stratified representation of the image. Note: The proposed multi-layer structure pyramid model differs from traditional hierarchical LDA model [15]. Level LDA model is the topic layer formed on the same vocabulary; by contrast, multi-layer structure pyramid LDA model is fixed by recursively formed words, it promotes a single variable of LDA model potential topics to change into Bayesian network variables. Here, we work on level LDA model with meticulous elaboration. Hierarchical LDA model is an extension of the LDA model, the theme depends on the hierarchy. Assuming a given layer L tree, each node is assigned a topic, the text will produce it in the following ways: 1) Choose a pate in the originality of the tree to a leaf node; 2) obtained vector from the L-dimensional Dirichlet in proportion theme; 3) along this route, words resulting from the mixing ratio of the document theme . Finally, the use of CRP (Chinese restaurant process) is to slack assumptions of hybrid tree structure. CRP is an integer division in the distribution, and can be imagined as a process in which the customers of amount M sitting at a Chinese restaurant with limit table number. Main process: the first customer sitting in a seat, then the m-th queuing customer seated in the following probability distribution.

$$p(OT_i \mid PC) = \frac{m_i}{\gamma + m - 1} \qquad (1)$$

$$p(NUT \mid PC) = \frac{m_i}{\gamma + m - 1} \qquad (2)$$

where OT_i refers to table i already occupied; NUT refers to the next unoccupied table; PC refers to customers already has a seat; m_i refers to the number of customers at the table. After M customers sit, the seat design would do implementation of the M division, this distribution and Dirichlet process [16] has the same partition structure.

Below, this paper proposes LDA-based hierarchical representation of the image. Compared to the previous latent factor model method, this paper is based on the potential underlying theme, formal hierarchical spatial distribution of an advanced method. In order to clearly explain the problem, first of all, this model is divided into two layers: L_0 and L_1, shown in Figure 1(a) above. By way of example, we can see in Fig 1(a), the example, L_0 layer structure of spatial grid size is 4×4, the gradient direction box size is of $V=8$ distribution of a word. Because the vocabulary of words corresponding to the position box, so that their occurrences refer to the histogram of energy, namely corresponding to the SIFT descriptors box.

In Figure 1(a) in, T_0 component mixing model is gained from a number of parameters $\phi_0 \in R^{T_0 \times X_0 \times V}$ space. In this example, $\phi_0 \in R^{T_0 \times (4 \times 4) \times 8}$. Integration of the L_1 layer includes the mixing ratio of L_0 layer, wherein, L_0 layer in the spatial characteristics of the grid from the L_1 region to X_1. L_1 constructs a component in the L_0 spatial distribution. T_1 component mixture model on the L_1 layer is parameterized via multiple parameters.

Spatial grid at each level and each observed words position of variables play a decisive role. However, the

distribution of words on the grid is not uniform, and the different components may differ. This must be calculated in a ϕ mixed distribution introducing X space (number) distribution on each level; all these spatial distributions are at different levels. Therefore, we need to define a complete derivative model. Note: The network size is 1×1 single-layer model and the model is equivalent to the LDA, and is a special case of the proposed method.

(a)

(b)
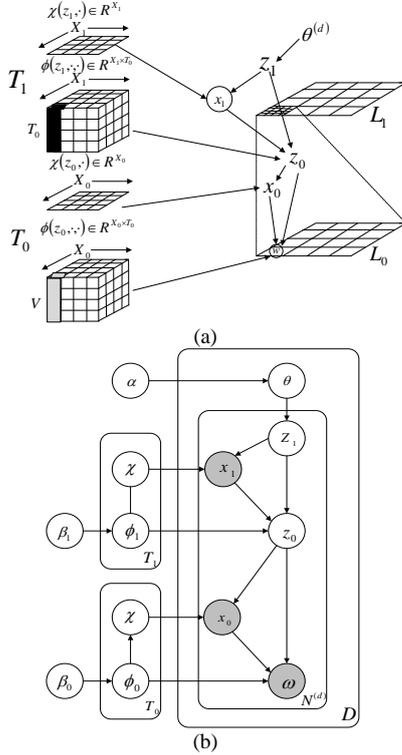
Figure 1. (a) Concept for the pyramid multilayer structure LDA model (b) Graphical model describing two-layers

A. Derivative Modeling Process

Figure 1(b) shows, it is a double pyramid LDA graphical model, variables are represented in the figure:

α, β_0 and β_1 are symmetric Dirichlet and priorities;

T_0 and T_1 are respectively number of mixture components corresponding to L_0 and L_1 layer;

$\phi_0 \in R^{T_0 \times X_0 \times V}$ and $\phi_1 \in R^{T_1 \times X_1 \times T_0}$ is introduced at each level space (number) distribution.

x_1, x_2, z_1, z_0 and w are in the following derivative and process 3) shows the definition.

α, β_0 and β_1 hyper parameters are given, the joint distribution of the model parameters can be broken down as follows:

$$p(w, z_{0,1}, \phi_{0,1}, x_{0,1}, \chi_{0,1}, \theta \mid \alpha, \beta_{0,1}) = p_{\phi_0} p_{\phi_1} \prod_{d=1}^{D} p_d \quad (3)$$

where,

$$p_{\phi_1} = \prod_{t_i=1}^{T_1} p(\phi_i^{(t_i,\cdot,\cdot)} \mid \beta_i) p(\chi_i^{(t_i,\cdot)} \mid \phi_i^{(t_i,\cdot,\cdot)})$$

$$p_d = p(\theta^{(d)} \mid \alpha) \prod_{n=1}^{N^{(d)}} p_{z_1} p_{z_0} p(w^{(d,n)} \mid \phi_0^{(z_0,x_0,\cdot)})$$

$$p_{z_1} = p(z_1^{(d,n)} \mid \theta^{(d)}) p(x_1^{(d,n)} \mid \chi_1^{(z_0,\cdot)})$$

$$p_{z_0} = p(z_0^{(d,n)} \mid \phi_1^{(z_1,x_1,\cdot)}) p(x_0^{(d,n)} \mid \chi_0^{(z_0,\cdot)})$$

In the above formula, parentheses superscript index is specific to each variable. "." refers to the whole range of the variable. For example, $\phi_i^{(t_i,\cdot,\cdot)}$ it refers to the polynomial parameters in grid and the underlying theme t_1 on topic L_0.

B. Learning and Reasoning

For the learning process of the model parameters, by inference mixture distribution is obtained for each word $w(d, n)$ and potential distribution $z_0(d, n)$ and $z_1(d, n)$ samples at the L_1 and L_0. Further, in FIG. 1(a), the observed position variables $x_0^{(d, n)}$, and $x_1^{(d, n)}$ through the grid X_0 and X_1 to track each occurrence of vocabulary.

Here, we briefly explain the Gibbs sampling method [14, 17]. As the single realization form of MCMC (Markov Chain Monte Carlo), its purpose is to construct a certain convergence of the probability distribution of the target. Sampling from the Markov chain is considered to be close to the probability. Objective probability distribution function is essential in the use of Gibbs sampling. Using Gibbs sampling, it's based on sampling current values and documentation set, to choose all variants value, from their distribution and then make transition to the next - state. Gibbs algorithm is as follows:

With y_i refers to a random number of $[1, T]$, T is the number of topics. $i \in [1, 2, \cdots, N]$, N is all the number of words occurrences of the text corpus, relating to the vocabulary size and vocabulary appear location. This is the initial state Markov chain;

According to the following formula, we will obtain the next state of the Markov chain, and carry enough iteration until Markov chain tends to target distribution, and marks y_i as current value.

$$p(y \mid w) = \left(\frac{\Gamma(w\beta)}{\Gamma(\beta)} \right)^T \sum_{j=1}^{T} \frac{\prod_w \Gamma\left(n_j^{(w)} + \beta\right)}{n_j^{(\cdot)} + w\beta} \quad (4)$$

Here, y is the subject which the word w is from; $n_j^{(w)}$ is the frequency to word j form w; $n_j^{(\cdot)}$ is the number of all words assigned topic j.

As for each sample y_i, we assess the ϕ and θ

$$\hat{\varphi}_j^{(w)} = \frac{n_j^{(w)} + \beta}{n_j^{(\cdot)} + w\beta} \quad (5)$$

$$\hat{\varphi}^{(d)} = \frac{n_j^{(d)} + \alpha}{n_*^{(d)} + T\alpha} \quad (6)$$

Where $n_j^{(d)}$ is the j is number of words assigned to the subject document j; $n_*^{(d)}$ is the number of all words assigned to topic j.

By Gibbs sampling inference methods, based on the model parameters joint distribution decomposition formula (3) and LDA theoretical model [13], put polynomial parametric model into integration, such as the push to equation (4). In equation (4), since all the variables x is observed, and between the ϕ and X, the conversion may be a decisive, therefore, can eliminate all of the terms in addition to terms involving x. Then, according to the following basic LDA, we obtain the equation (4) results:

$$p(\theta, z \mid w, \alpha, \beta) = \frac{p(\theta, z, w \mid \alpha, \beta)}{p(w \mid \alpha, \beta)} \qquad (7)$$

$$p(\theta, z, w \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta)$$

$$p(w \mid \alpha, \beta) \int p(\theta \mid \alpha) \left(\prod_{n=1}^{N} \sum_{z_n} p(z_n \mid \theta) p(w_n \mid z_n, \beta) \right) d\theta$$

$$p(\theta \mid \alpha) = \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}$$

$$\theta_i \geq 0, \quad \sum_{i=1}^{k} \theta_i = 1$$

where, N is the number of words in a text; θ can be seen as a random variables of k-dimensional vector, parameters α component $\alpha_i > 0$. Between different layers through the variable x will perform spatial grouping. The formula (7) is easy to generate L-layer model, as shown in Equation (8). Note that the formula (8) in the "evidence" refers to the famous Bayesian evidence:

$$p(w, z_{0,1} \mid x_{0,1}, \alpha, \beta_{0,1}) = \int_\theta \int_{\phi_1} \int_{\phi_0} \int_{\chi_1} \int_{\chi_0} p(w, z_1, z_0, \phi_1, \phi_0, \theta, \chi_1, \chi_0 \mid x_1, x_0, \alpha, \beta_1 \beta_0) d\chi_0 d\chi_1 d\phi_0 d\phi_1 d\theta$$

$$= \int_\theta \int_{\phi_1} \int_{\phi_0} p(w, z_1, z_0, \phi_1, \phi_0, \theta \mid x_1, x_0, \alpha, \beta_1, \beta_0) d\phi_0 d\phi_1 d\theta \qquad (8)$$

$$= \int_\theta p(\theta \mid \alpha) p(z_1 \mid \theta) d\theta \int_\phi p(\phi_1 \mid \beta_1) p(z_0 \mid \phi_1, z_1, x_1) d\phi_1 \int_\phi p(w \mid \phi_0, z_0, x_0) p(\phi_0 \mid \beta) d\phi_0$$

$$p(w, z_{1,\cdots,L} \mid x_{1,\cdots,L}, \alpha, \beta_{1,\cdots,L}) =$$
$$= \int_\theta p(\theta \mid \alpha) p(z_L \mid \theta) d\theta \prod_{l=1}^{L-1} \int_{\phi_l} p(\phi_l \mid \beta_l) p(z_{l-1} \mid \phi_l, z_l, x_l) d\phi_l \int_\phi p(w \mid \phi_0, z_0, x_0) p(\phi_0 \mid \beta) d\phi_0 \qquad (9)$$

III. PERFORMANCE EVALUATION

To evaluate the performance data of proposed probabilistic model multilayer structure pyramid, this section tests the Caltech 101 data sets [18]. Set contains 101 classified objects. Different classes of objects contain different number of images. Experimental has been compared the performance of each component of a different number, and explore a single model, feed forward LDA model (FLDA) and the proposed total probability generation model (RLDA). The results showed that:

1) In the monolayer model adding a layer may improve the classification performance;

2) RLDA feed forward model improves performance of the FLDA model.

Implementation

This paper presents methods of image feature representation is 20×20 pixels SIFT descriptors, which is spanned from 8 pixels to accurately extract pictures. Each descriptor is processed by proposed probabilistic model. Since LDA model requires that the count data is discrete, and the SIFT dimension is continuous values, so the maximum value normalized to 200 SIFT mark, this quantization level descriptors retains enough information.

Experimental compares the following three models:

LDA: LDA Model with 30K SIFT feature extraction area, training different number of components (258, 512 and 1024). In addition, "superior region" is contained of 8×8 SIFT feature area also trained LDA model.

FLDA: The feed forward model is the first region in the training SIFT features to practice an LDA model. LDA model study on the input side, integrating 8×8 SIFTS" superior region" as the output. Experiment tested the 512 at the bottom components and tested 512 and 2048 components at the top level.

Evaluation

Classification test follows spatial pyramid matching model, and it is general method [3] of the Caltech-101 Evaluation . A spatial pyramid has three layers: contains 8×8, 4×4, 2×2 grids, which is built on top of features. Using the maximum pool [6] to pyramid polymerization space and using linear SVM to classification. Libsvm is selected as experimental SVM training lab environment. 30 for each image class object is used for training, the remaining image for testing. Experiment runs 20 times to obtain average accuracy and standard deviation. Experiments, Gibbs sampling topics mesh T = 100, hyper parameters α =60/T, $\beta_{0,1}$ =0.02. Adopting proposed image hierarchical representation model for the entire modeled theme of document set , Diego generation of 2000 times, each text is represented as a main theme that contains 200 multinomial distribution on the set of questions, getting documentation set implicit theme in the constructed on a linear SVM classifier.

As for the comparison model, LDA model calculation complexity is $O(N^2 k)$, and N is the number of words in a text, k is random variable dimension. FLDA model core implementation is still LDA model, but using a feed forward manner, so the calculation of complex is the same as LDA model. Hierarchical LDA model was

constructed as a hierarchical tree, therefore, the calculation complexity is $O(N^2k)$, As for RLDA model, because it is pyramid multilayer structure, an increase for each additional layer is equivalent to a extra layer of the tree, so its computational complexity is the same as level of LDA . In fact, based on the models constructed, linear SVM classifier aims to predicate the experimental data classification. Since we use the linear SVM classifier device, and the purpose of our experiment is to classify the data set correctly, therefore, based on the output data of the model are linearly SVM classification, the classification of the computational complexity is also a linear relationship.

First, test LDA model on the single SIFT features ,so with 128 components obtained classification accuracy of 58.4%, weight increase to 512, the classification accuracy increased to 69.6%, component increased to 1024, the classification accuracy of 71.3%.

Here, the focus on the evaluation of the impact of the increase of presentation layer on classification performance. In this test, the number of components is of a given topic is 512, training and comparing with three models described in Chapter II.A: single LDA, double FLDA and RLDA model. Single model and the two-layer model for the sub-class is the same dimension of feature vectors. Double bottom and top model layer of 512 components for training, were recorded as 512b and 512t, and they correspond to the SIFT descriptor. Table 1 shows the components of the experimental results of 512 models in L_1 layer and 512 components modes in $L2$ layer. Here, "bottom" refers to the area in the SIFT feature training, "top" means the 8×8 SIFT's "Parent Zone" on the train, "both" refers to the superposition trained on two characteristics.

TABLE I.   COMPARISON OF EXPERIMENTAL RESULTS OF THREE MODELS

Methods				Caltech-101
	Model	Size	Use	30 training sets
128 Dimensions Model	LDA	512	bottom	58.7 $\pm$ 2.1%
	RLDA	512b/512t	bottom	63.4 $\pm$ 0.7%
	LDA	512	top	61.5 $\pm$ 1.0%
	FLDA	512b/512t	top	60.3 $\pm$ 1.3%
	RLDA	512b/512t	Top	65.0 $\pm$ 1.2%
	FLDA	512b/512t	both	65.2 $\pm$ 1.0%
	RLDA	512b/512t	Both	69.3 $\pm$ 0.7%

As it is evident from Table 1, in the area of training the RLDA model of SIFT features, the classification accuracy of 61.6%, more than the 4.9 percentage points compared with single-LDA model classification accuracy 56.9%. This indicates that joint learning has outstanding performance than the traditional single-LDA model. In addition, the two-layer model of the single-layer model to improve the classification performance, that FLDA correct classification rate of 62.3. Both SIFT feature greatly improves the classification accuracy was 67.0%. Regional monolayer model accuracy is 58.9% when tested on the classification, as well as in 8×8 SIFT "upper

zone" monolayer model when tested on the classification of fine degree of 61.5%.

Figure 2 (a) shows the number of training samples of average accuracy rate at different image points. Apparently, RLDA classification performance of the model is always superior to FLDA model, FLDA classification performance of the model. In addition, experimental tests of the L_1 and L_2 layers superimposed onto a characteristic feature vector under the classification performance (for example, Table 1, "both" case), this feature includes more information, taking into account both of the characteristics of the larger space, also considering a small special sign. Figure 2 (b) shows the performance comparison, when use of this information of the "both" case, FLDA is further improved by nearly 4% accuracy, RLDA increased by 3% accuracy.

Figure 2 Comparison of classification with different number of training samples.
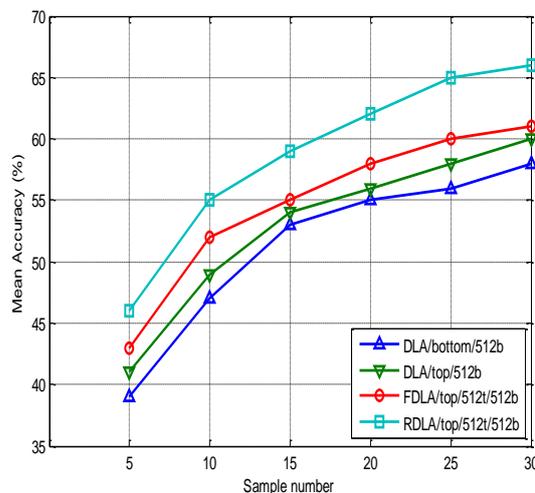
(a) "bottom" and "top" case
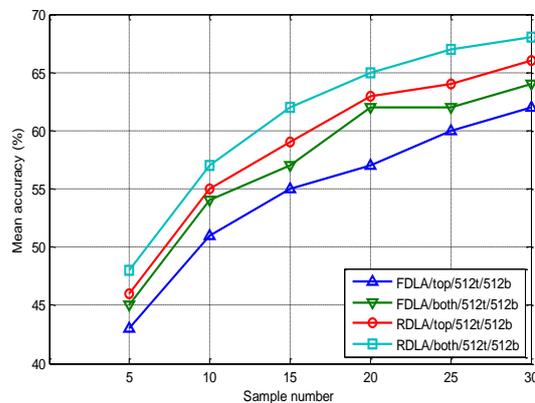
(b) "both" case

Figure 2.   Comparison of classification accuracies with different number of training samples on Caltech-101 dataset

Next, the 1024 top components and 128 bottom components, learning on the two-layer mode. It was found that these models classification performance's difference is smaller than before, but it is higher than 128b/128t case's classification performance. Specifically, degree of RLDA model study on the top floor was 71.5%, while the FLDA model learning precision is 72.5% and

monolayer model classification accuracy is 67.8%. RLDA model at the bottom learning accuracy up to 63.7%, FLDA model reached 63.6% meanwhile. RLDA reached 74.7%, FLDA's accuracy, when using double, reached 73.9%.In addition to the above experiments, in this section we will present model associated with the hierarchy model (Hierarchical-LDA [15], CNN [19], CNN + Transfer [2], CDBN [20] and Hierarchy-of-part [20] Model), respectively with 20 and 35 images of the training set images for linear SVM classifier performance for ratio comparison on the Caltech-101 data sets, as shown in Tables II and III. From the experimental results in Table II clearly shows, most existing hierarchical model outperforms the classification SIFT features single test area classification accuracy of LDA model 58.9%; while the proposed RLDA model hierarchy is seven percentage point higher than the existing model type of classification accuracy.

TABLE II. COMPARISON OF CLASSIFICATION ACCURACIES BETWEEN RLDA MODEL AND OTHER HIERARCHICAL MODELS BY USING 30 TRAINING IMAGES

Methods			Caltech-101
Models		Layer user	30 training set
Proposed model	RLDA(1024t/128b) RLDA(1024t/128b) RLDA(1024t/128b)	bottom top Both	$63.7 \pm 0.5\%$ $73.6 \pm 1.2\%$ $74.7 \pm 0.8\%$
Layer model	Hierarchical-LDA CNN CNN CNN+Transfer CDBN Hierarchy-of-parts	Both bottom top top both both	$68.4 \pm 0.7\%$ $58.6 \pm 0.4\%$ $67.3 \pm 1.5\%$ 68.2% $66.4 \pm 0.4\%$ 67.5%

TABLE III. COMPARISON OF CLASSIFICATION ACCURACIES BETWEEN RLDA MODEL AND OTHER HIERARCHICAL MODELS BY USING 15 TRAINING IMAGES

Methods			Caltech-101
Models		Layer use	15 taining set
Proposed model	RLDA(1024t/128b) RLDA(1024t/128b) RLDA(1024t/128b)	bottom top both	$57.6 \pm 0.8\%$ $67.7 \pm 0.9\%$ $68.4 \pm 0.5\%$
Layer model	Hierarchical-LDA CNN CNN CNN+Transfer CDBN Hierarchy-of-parts	both bottom top top both both	$63.1 \pm 0.8\%$ -- -- 59.1% $58 \pm 1.5\%$ 61.5%

Currently, most of the existing hierarchical models are presented with the best performance of layer model learning, and in the literature [11, 12, 20], the experimental results showed that they added an identification layer but decreased fine degrees; while in the proposed Bayesian model, it can be seen from the experimental results, adding a presentation layer enhances the classification performance. Most importantly, full Bayesian model proposed in image classification outperforms former feed-forward method. This indicates that Bayesian inference resulted in far

more stable assessment of the regional characteristics, and has been robust recognition performance. Hierarchical recognition highlights the importance of feedback.

In addition, Figure 3 presents the total probability feed forward model and generate models learning component of the feature region in the image comparison visualization. FLDA only in the local edge direction for learning, and RLDA learning is more complex spatial structure. As can be seen by Figure 3, FLDA feed forward model, the underlying theme is essentially a direction. The second layer is not attached on them any other structure. Thus, there appears the top topic in the same direction which is limited to a space supported by a larger range; while as for RLDA model, the top component seems to exhibit more potential, discriminative space frame structure. Meanwhile, we can also find the characteristics of the underlying neighborhood RLDA, within the subject there is a strong relationship between the activities, indicating that the mold type of reasoning can be rendered by the underlying connection between the sub-space continued structures.

(a)　　　　(b)
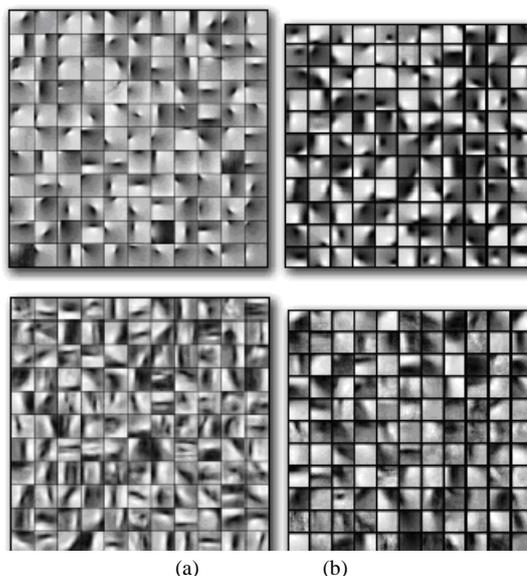
Figure 3. Comparison of the (a) feed-forward FDLA and (b) entire yielded models

IV. CONCLUSIONS

For higher complexity visual features, we proposed a joint training of all presentation layer probability models. This model has better classification performance than the existing hierarchical model. In addition, it indicates by adding additional layers significantly to improve the classification performance, and this method has better performance than the method for feed forward. This shows the importance of the hierarchical model of visual recognition of feedback. In conclusion, the proposed probabilistic model is robust, easy to modular combination. With the increase of layers, this model can be characterized from a source to all intermediate object recognition for methods to evolve.

REFERENCES

[1] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV),* vol. 60, no. 2, pp. 91-110, 2004.

[2] A. Ahmed, K. Yu, W. Xu, Y. Gong, E. P. Xing. Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks. *Proceedings of the 10th European Conference on Computer Vision* (ECCV): Part III, Marseille, France, Berlin: Springer-Verlag, pp. 69-82, 2008.

[3] S. Lazebnik, C. Schmid, J. Ponce. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), NY, USA: *IEEE Computer Society*, pp. 2169-2178, 2006.

[4] J. Yang, K. Yu, Y. Gong, T. Huang. Linear spatial pyramid matching using sparse coding for image classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Florida, *USA: IEEE Computer Society*, pp. 1794-1801, 2009.

[5] B. A. Olshausen, D. J. Field. Sparse coding with an over-complete basis set: a strategy employed by V1?. *Vision Research*, vol. 37, no. 23, pp. 3311-3325, 1997.

[6] Y. L. Boureau, F. Bach, Y. LeCun, J. Ponce. Learning mid-level features for recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, USA: *IEEE Computer Society*, pp. 2559-2566, 2010.

[7] M. Fritz, M. J. Black, G. R. Bradski, S. Karayev, T. Darrell. An additive latent feature model for transparent object recognition. Proceedings of the 23rd Annual Conference on Neural Information Processing Systems (NIPS), Vancouver, B.C., Canada: MIT Press, pp. 558-566, 2009.

[8] J. Mutch, D. G. Lowe. Object class recognition and localization using sparse features with limited receptive fields. International Journal of Computer Vision (IJCV), vol. 80, no. 1, pp. 45-47, 2008.

[9] E. T. Rolls, G. Deco. Computational neuroscience of vision. Oxford University Press, 2002.

[10] H. Lee, R. Grosse, R. Ranganath, A. Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Proceedings of the 26th Annual International Conference on Machine Learning (ICML), NY, USA, New York: ACM, pp. 609-616, 2009.

[11] M. A. Ranzato, F. J. Huang, Y. L. Boureau, Y. LeCun. Unsupervised learning of invariant feature hierarchies with applications to object recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Alaska, USA: IEEE Computer Society, pp. 1-8, 2008.

[12] T. Serre, L. Wolf, S. Bileschi, T. Poggio. Object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 3, pp. 411-426, 2007.

[13] J. Sivic, B. C. Russell, A. A. Efros. Discovering objects and their locations in images. Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV), Beijing, China, New York: Springer-Verlag, pp. 370-377, 2005.

[14] D. M. Blei. Latent Dirichlet allocation. Journal of Machine Learning Research, vol. 3, no. 1, pp. 993-1022, 2003.

[15] T. S. Lee, D. Mumford. Hierarchical Bayesian inference in the visual cortex. Journal of the Optical Society of America Association, vol. 20, no. 7, pp. 1434-1448, 2003.

[16] D. M. Blei. Hierarchical topic models and the nested Chinese restaurant process. Advances in Neural Information Processing Systems (NIPS) 16, British Columbia, Canada: MIT Press, 2010.

[17] T. Ferguson. A Bayesian analysis of some nonparametric problems. The Annals of Statistics, vol. 1, no. 3, pp. 209-230, 1973.

[18] G. Heinrich. Parameter estimation for text analysis. Technical Report, Darmstadt, 2008.

[19] L. F. Fei. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. Journal of Computer Vision and Image Understanding, vol. 106, no. 1, pp. 59-70, 2004.

[20] K. Kavukcuoglu. Learning convolutional feature hierarchies for visual recognition. Advances in Neural Information Processing Systems (NIPS), Vancouver, B.C., Canada: MIT Press, pp. 1090-1098, 2010.

[21] S. Fidler. Similarity-based cross-layered hierarchical representation for object categorization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Alaska, USA: *IEEE Computer Society,* pp. 1-8, 2008.

Fushun WANG was born in Hebei province of China at October 10, 1981. He received his bachelor degree and master degree from College of Mechanical and Electrical Engineering, Hebei Agricultural University, China, in 2003 and 2006 respectively. Currently, he is a lecturer of College of Information Science and Technology, Hebei Agricultural University, China. His research interests include Computer application, Computer monitoring and control , Computer networks and distributed systems.

Enhancement Contrast and Denoising of Low Illumination Image of Underground Mine Tunnel

Wanfeng Shang, Hongwei Ma, and Xuhui Zhang

College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, P. R. China

Email: shangwanfeng@gmail.com, {mahw, zhangxh}@xust.edu.cn

Abstract—For enhancing low lumination image of underground mine tunnel, a hybrid local histogram equalization with partial differential equation is proposed to combine a local histogram equalization and total variation method with edge-stopping function via partial differential equation. The local histogram equalization is to enhance contrast of image with keeping shape and edge, while total variation method with edge-stopping function is to smooth image by gradient diffusion for lowing noise. In the presented hybrid method, a parallel trade-off scheme is utilized for simultaneous denoising and contrast enhancement. Experiments are carried out to validate the hybrid method using a standard image and a real image of underground mine tunnel. The results prove that it is an alternative processing method for low luminous image of underground mine tunnel.

Index Terms—Histogram Equalization, Denoising, Contrast Enhancement, Image Processing, Total Variation Method

I. INTRODUCTION

The environment in underground coal mine is very complex, especially after an accident [1]. Robots have a great potential to assist in the underground operation, searching ahead of rescue teams and reporting conditions that may be hazardous to the teams by providing video and atmospheric monitoring information at several emergency and recovery sites [2, 3]. As one of the key technologies image processing technology has been paying attention in visual sensation with low illumination of autonomous mine rescue robot. Underground coal mine environment is low illumination and more dust, making the camera lens covered with a layer of dust, thus affecting the quality of image. The acquired image often have low contrast and more noise, blurring the details of the original characteristics of the image and making human visual resolution or machine identification more difficult. Therefore, it is necessary to preprocess image for further control of coal mine rescue robot, mainly consisting of denoising and contrast enhancement of these images.

Image enhancement is one of image process technologies, not considered the reasons of low quality of image, only highlighting interested features in image and attenuating unwanted characteristics. The improved image is not necessarily close to the original image, by projecting object contour, denoising, and strengthening contrast. From the viewpoint of the image quality evaluation, the image enhancement technology is to make image more suitable than the original image for human visual or machine recognition. Many image enhancement methods were proposed to enhance images degraded by low illumination: linear contrast stretching, global histogram equalization and local histogram equalization etc. [4]. They are simple transformations for contrast enhancement, but they do not always produce good results. Particularly for images with large spatial variation in contrast, they causes over enhancement contrast in some portion of the image and undesired effect of more noise in the input image, along with the image features, further weakening minutia part. To improve the process of denoising, researchers developed frequency-based filters [5-7], average within neighborhood region [8-10], and partial differential equation [11, 12], etc.. A low pass filter or average filtering can smooth the noise image, but blurring its details, because both noise and edge in image are within high frequency domain [13]. Discrete cosine and wavelet transform methods are used to alter the frequency content of an image to improve desired traits for low complexity of computations and conveniently manipulating the frequency composition of the image.[5] However, they still have some basic limitations, for example, introducing block artifacts. Partial differential equations (PDE)-based filters are modeling an image restoration or enhancement process through a partial differential equation that regards a degraded image as the initial state of a diffusion process and relates the image's spatial derivatives with its time derivative. A large number of PDE-based methods have been proposed to tackle the problem of image denoising. The regularized P-M method [11] smoothed the edges of the image slowly and the edges are kept, but some details still lost. Basing on the above analysis, we can find that enhancing edges in image will simultaneously increasing noise in the image, in turn, reacting on the edge to be fuzzy to a certain extent. Therefore, in the paper, we propose a compromise method to for both denoising and contrast enhancement of image.

The paper presents an integrated enhancement method of PDE-based filter and local histogram equalization. The local histogram equalization is to enhance contrast of

connected component in image for preserving detail shape, while regularized diffusion function along the tangent direction of the edge serves as a filter. In addition, it is important reason that one of the advantages of the use of PDE for image processing is the more possibility to combine algorithms. In general, there are two combination schemes. One is to reduce noise before stretching contrast of image, the other is to stretch contrast in first, and then denoise of image. However, we adopt a parallel trade-off scheme for the proposed hybrid local equalization with PDE (HLEPDE) for simultaneous denoising and contrast enhancement.

This paper is organized as follows. Section 2 presents the proposed PDE-based histogram equalization for image enhancement. In section 3, simulations are carried out to validate the proposed method by processing a test image and an underground image in coal mine. Finally, a conclusion is drawn in section 4.

II. PDE BASED HYBRID ENHANCEMENT METHOD OF IMAGE

A. Global Histogram Equalization based on PDE

Global histogram equalization (GHE) [4] is one of the basic and most useful operations in image processing, improving image quality by extending dynamic range of intensity using the histogram of the whole image.

Let I be an image defined in $N \times M$ size with grey values in the range [a, b], where a and b are respectively minimal and maximal value of image grey. For continual image, we can first define an area function

$$A(D) := Area\{(x,y) : I(x,y) \geq D\} \quad (1)$$

It displays overall area of image $I(x,y)$ not less than the given threshold D. When D increasing to $D+\triangle D$,

$$A(D+\Delta D) \approx A(D) + \frac{dA(D)}{dD}\Delta D \quad (2)$$

and then the difference between $A(D+\triangle D)$ and $A(D)$ is area of image in range $[D, D+\triangle D]$. So we can get the following form

$$-\frac{dA(D)}{dD} \approx \frac{Area\{(x,y), D \leq I(x,y) \leq D+\Delta D\}}{\Delta D} \quad (3)$$

When considering $\triangle D$=1 in a digital image, the right of Eq. (3) is sum of pixel of grey level equal to D in image. While for continual image, the left of Eq. (3) must divide the overall image A_Ω, corresponding to histogram definition of digital image. Hence, the histogram of continual image can be expressed as

$$h(D) = -\frac{dA(D)}{A_\Omega dD} \quad (4)$$

As Eq. (4) is showed, $A(D)$ is decreasing, leading to d$A(D)$/d(D)<0. It can derive that the histogram $h(D)$ of continual image must be positive. The cumulative distribution function of continual image is

$$H(D) = \int_0^D h(\xi)d\xi \quad (5)$$

subject to $H(a) = 0$ and $H(b) = 1$

The histogram equalization corresponds to selecting $h(\bullet)$ to be the distribution function $H(\bullet)$ of I. Basing on Eq.(1)-Eq.(5), $H(\bullet)$ is given by

$$H(D) = \frac{A_\Omega - A(D)}{A_\Omega} = \frac{Area\{(x,y) : I(x,y) \geq D\}}{A_\Omega} \quad (6)$$

Obviously, $H(\bullet)$ is strictly increasing. The transformation function for histogram equalization image is

$$f(D) = (b-a)H(D) + a \quad (7)$$

and it is a monotonic increasing function. Hence, for every grey value of pixel in input image I_A, there is a corresponding output using the relation

$$I_B(x,y) = f(I_A(x,y)) \quad (8)$$

In addition, it follows that the basic information of an original image is contained in the family of level sets

$$\chi_{D_A}[I_A] = \{(x,y) \in \Omega : I_A(x,y) \geq D_A\} \quad (9)$$

for all values of D_A in the range of I_A. Observe that, under fairly general conditions, an image can be reconstructed form its level sets by the formula

$$I_A(x,y) = \sup\{D_A : (x,y) \in \chi_{D_A}[I_A]\} \quad (10)$$

Since $H(\bullet)$ is an increasing function, the linear transformation of Eq.(8) does not modify the family of level-sets of I_A. We can get

$$\chi_{D_B}[I_B] = \chi_{D_A}[I_A] = \{(x,y) \in \Omega : I_A(x,y) \geq D_A\} \quad (11)$$

Therefore, the global histogram equalization of image basing on Eq.(8) is a homeomorphic transformation.

Further, a resulting image by histogram equalization can be expressed $I(x,y,t)$ as the following equation developing

$$\frac{\partial I(x,y,t)}{\partial t} = [1 - \frac{I(x,y,t) - a}{b - a}]A_\Omega - A(I(x,y,t)) \quad (12)$$

subject to $I(x,y,0) = I_0(x,y)$

It was shown in [14] that the equalization of an image was to minimize the function

$$E(I) = \frac{\Omega}{2(b-a)} \int_\Omega (I(x,y) - \frac{b-a}{2})^2 dxdy$$
$$-\frac{1}{4}\int_\Omega \int_\Omega |I(x,y) - I(u,v)| dxdydudv \quad (13)$$

where the first term tries to keep the grey values of I near as near as possible to the mean $(b-a)/2$ and the second term is a measure of the contrast of the whole image. Obviously, Eq. (12) is gradient descent flow of Eq.(13) and it has only stable resolution when is equal to zero.

$$\frac{\partial I(x,y,t)}{\partial t}=0 \Rightarrow$$

$$I(x,y,\infty)=\frac{A_{\Omega}-A(D)}{A_{\Omega}}(b-a)+a=(b-a)H(D)+a \quad (14)$$

The resolution is same with Eq. (13). When Eq. (14) is up to be stable, the image I has finished the global equalization. So Eq. (12) for histogram equalization is feasible and stably convergent and is rewritten using Eq. (7) and Eq. (14).

$$\frac{\partial I(x,y,t)}{\partial t}=(\frac{A_{\Omega}}{b-a})(\frac{A_{\Omega}-A(I(x,y,t))}{A_{\Omega}}(b-a)+a-I(x,y,t))$$

$$=(\frac{A_{\Omega}}{b-a})(f(I(x,y,t))-I(x,y,t)) \quad (15)$$

In order to restrict change of grey value within the range [-255 255], the above equation is rescaled by multiplying constant coefficient $(b-a)/A_{\Omega}$. Hence, the general expression of histogram equalization based on PDE is shown by

$$\frac{\partial I(x,y,t)}{\partial t}=f(I(x,y,t))-I(x,y,t) \quad (16)$$

B. Shape Preserving Local Histogram Equalization

A considerable large amount of the research in image processing is based on assuming that regions with almost equal grey values, which are topologically connected, belong to the same physical object in the word. Following this it is natural to assume then that the shapes in a given image are represented by its level sets. If the relation between the original and resulting image is not homeomorphic using some local contrast enhancement methods, these methods may be undesirable for extracting and recognizing feature. This paper introduces a local histogram equalization, called shape preserving local histogram equalization (SPLHE) [12], to modify the transformation function $f(\bullet)$ in Eq.(11) for enhancing contrast and preserving the shape of image.
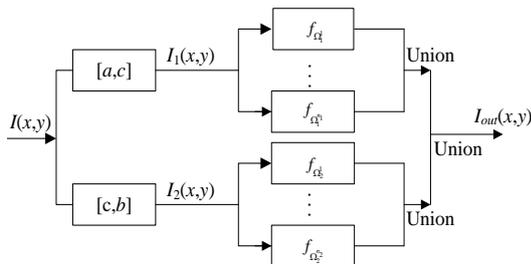
Figure 1.   Local histogram equalization scheme

Suppose pixels of I in the domain $\mathbf{\Omega}$ have grey value in range [a, b]. Now divide I into two subdomains and get two sets.

$$\Omega_1 := \{(x,y), a \le I(x,y) \le c\} \\ \Omega_2 := \{(x,y), c \le I(x,y) \le b\} \quad (17)$$

subject to

$$\Omega_1 \cap \Omega_2 = \varnothing \text{ and } \Omega_1 \cup \Omega_2 = \Omega \quad (18)$$

The edge of subdomain is a closed curve of part of level set $I(x,y)=c$. There are existing two subimages in these two subdomains, taking the following form:

$$I_1(x,y) = \begin{cases} I(x,y), & \forall (x,y) \in \Omega_1 \\ 0 & \text{others} \end{cases} \quad (19)$$

$$I_2(x,y) = \begin{cases} I(x,y), & \forall (x,y) \in \Omega_2 \\ 0 & \text{others} \end{cases} \quad (21)$$

In a similar way, if continuously dividing subimage into two sections, we get much more subdomains.

$$\Omega_1 = \cup\Omega_1^i, \quad \Omega_2 = \cup\Omega_2^i \quad (i=1,2,...,k; \ N \approx 2^{k+1}) \quad (21)$$

Here, Ω^i_1 and Ω^i_2 is respectively called a connected component of Ω_1 and Ω_2. A local image processing is to enhance connected component of image. The transformation function can be rewritten basing Eq.(7) for equalizing connected component of image because the global histogram equalizing is a homeomorphic transformation.

$$f_{\Omega}(I(x,y))=\sum_{i}^{k}f_{\Omega_1^i}(I_{\Omega_1^i}(x,y))+\sum_{i}^{k}f_{\Omega_2^i}(I_{\Omega_2^i}(x,y)) \quad (22)$$

where

$$\left. \begin{array}{l} f_{\Omega_1^i}(I_{\Omega_1^i}(x,y))=(c_i-a_i)H_{\Omega_1^i}(I_{\Omega_1^i}(x,y))+a_i \\ f_{\Omega_2^i}(I_{\Omega_2^i}(x,y))=(b_i-c_i)H_{\Omega_2^i}(I_{\Omega_2^i}(x,y))+c_i \end{array} \right\} \quad (23)$$

Therefore, Eq.(22) and Eq.(23) constitute a local equalization method basing on PDE, whose structure [15] shows in Fig. 1. The advantage of local histogram equalization is better contrast enhancement with preserving local shape, while the use of PDE for the local image process makes it possible to combination of different methods for better quality image. We will later see how to simultaneously reduce noise and enhance contrast of image.

C. Hybrid Enhancement Method via PDE

Local histogram equalization allows a better contrast enhancement with preserving local shape of image. The use of PDE method makes it possible to combine different methods for better quality image. A flow for simultaneous denoising and histogram modification is presented in this section.

A smooth operator can be achieved by minimizing the total variation of the image [12], given by

$$\min \int_{\Omega}|\nabla I(x,y)|\,dxdy \quad (24)$$

The Euler-Lagrange of this functional is given by the curvature of level-sets of image, which leads to the gradient descent flow

$$\frac{\partial I(x,y,t)}{\partial t}=\text{div}(\frac{\nabla I(x,y,t)}{|\nabla I(x,y,t)|}) \quad (25)$$

This flow smoothes the image in the parallel direction to the edges for preserving anisotropic diffusion. Using this smoothing operator, together with the histogram modification part, gives very similar results as those obtained with the affine based flow. If this smoothing operator is combined with the histogram flow, the total flow

$$\frac{\partial I(x,y,t)}{\partial t} = \lambda \mathrm{div}(\frac{\nabla I(x,y,t)}{|\nabla I(x,y,t)|}) + \qquad (26)$$
$$f(I(x,y,t)) - I(x,y,t)$$

will therefore be such that it minimizes

$$\lambda \int_{\Omega} |\nabla I(x,y)|\, dxdy + E(I) \qquad (27)$$

where λ is a positive parameter controlling the trade-off between smoothing and histogram modification, called balance factor, $E(I)$ is shown by Eq.(13). In order to reduce diffusion rate of near edge, Eq.(26) is modified using an edge-stopping function. Hence, the proposed HLEPDE is shown by

$$\frac{\partial I(x,y,t)}{\partial t} = \lambda g(|\nabla I(x,y,t)|)\mathrm{div}(\frac{\nabla I(x,y,t)}{|\nabla I(x,y,t)|}) \quad (28)$$
$$+ f(I(x,y,t)) - I(x,y,t)$$

where

$$g(r) = \frac{1}{1+(r/K)^2} \qquad (29)$$

The function $g(\bullet)$ becomes gradually larger with the diffusion from a local plane region to edge, and hence the first term in Eq. (28) gradually increases and the denoising and smooth effect will strengthen. The Eq. (28) via PDE function permits to consider simultaneous denoising and local histogram modification, and the edge-stopping function makes it to adaptively adjust diffusion rate for image processing. However, the I image must be preprocessed by Gaussian regularization because the noise in I image may provide false gradient information, leading to a unstable solution for Eq. (28).

For a digital image, the main strategy of the HLEPDE is that its PDE in a discrete form is performed iteratively, based on the implements of SPLHE to enhance contrast of edge and diffusion flow of total variation of the image to smooth edge, and then balance these two functions by a trade-off factor. Fig. 2 shows the outline procedure of the hybrid algorithm, and the detailed procedure involved in HLEPDE is described as follows.

Step 1: Initialize the parameters, set iteration number $n=1$.

Step 2: Gaussian regularization of the input image to prevent false edge occurring.

Step 3: Compute divergence information of the input image and then local histogram equalization of SPLHE is carried on the input image.

Step 4: The output image is gotten by using a discrete form of Eq. (28).

$$I_{n+1} = \lambda g(|\nabla I_n|)\mathrm{div}(\nabla I_n) + f(I_n) - I_n \qquad (30)$$
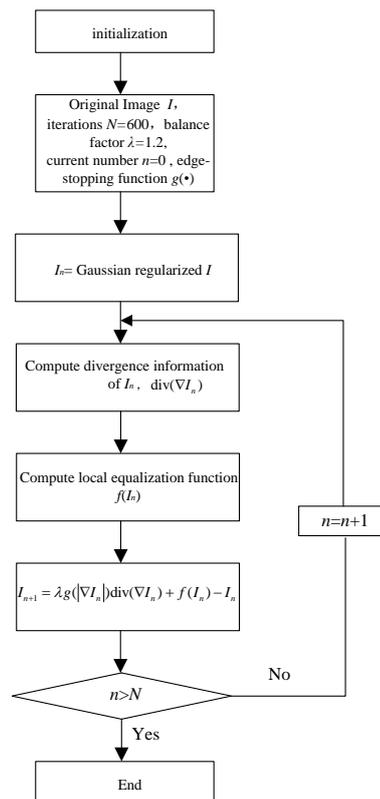
Figure 2.   Flow diagram of HLEPDE

Step 5: Termination judgment. The loop terminates if the number of current iteration n reaches the maximumiteration N. Otherwise, the program goes back to the step 3, n= n +1, and repeats executing the remaining steps.

Step 6: The program completes and obtains processed image with better quality.

III.   EXPERIMENTS

In this section, the proposed HLEPDE based on Fig. 2 is compared with other image processing methods to validate its effect by enhancing a standard image and practical image of underground mine tunnel.

A noised image is produced by adding random Gaussian noise to a 512×512 pixel standard test image, called Lenna. An original image added into gauss noise, shown in Fig. 3(a), is enhanced using different methods. Since Fig. 3(b) is resulted by only using SPLHE for contrast enhancement, there remains a lot of noise and even more.

However, only using TV [15] method for denoising has also bad effect, as shown by Fig. 3(c). So we incorporate the diffusion flow of TV method and local enhancement of SPLHE for complementary image processing. There are three ways to combine these two methods for making full of use of their advantages. One is to first stretch contrast using SPLHE method and then denoise with TV, by which we enhance Fig. 3(a) and get Fig. 3(d). On the contrary, the other way is to first utilize TV method then SPLHE, by which Fig. 3(e) is attained.

(a) Original image   (b) Image using SPLHE   (c) Image using TV

(d) Image using SPLHE before TV   (e) Image using SPLHE after TV   (f) Image using HLEPDE

Figure 3.   Image processing using different methods

TABLE I.   INDEX DATA OF TEST IMAGE USING DIFFERENT METHODS

	Mean intensity	Entropy	RMSE	PSNR	Average gradient
SPLHE	132.1163	7.4011	36.2310	16.9492	17.2582
TV	118.2512	7.5464	11.0485	27.2647	5.9030
SPLHE then TV	132.1324	7.8790	38.3591	16.4534	11.0172
TV then SPLHE	133.6356	7.3353	36.0948	16.9819	9.3244
HLEPDE	130.5254	7.9395	30.3357	18.4917	12.0784

as Fig. 3(d) and Fig. 3(e), they are still obvious block effect and lower contrast. In addition, the third way is the proposed method that is a parallel scheme for simultaneous denoising and contrast enhancement. Comparing Fig. 3(d), Fig. 3(e) and Fig. 3(f), we achieve that Fig. 3(f) has higher contrast, less noise, and more distinct.

In order to further evaluate the performance of the proposed, we introduce some characteristic indices. That is mean intensity (MI), standard deviation (SD), entropy, RMSE (root mean square error), peak signal-to-noise ratio (PSNR), average gradient (AG) of image, as defined as the following equations.

$$MI = \frac{1}{N*M}\sum_{x=1}^{N}\sum_{y=1}^{M}I(x,y) \qquad (31)$$

$$Entropy = -\sum_{g=0}^{255}P_g \log_2 P_g \qquad (32)$$

$$RMSE = \sqrt{\frac{1}{M*N}\sum_{x=1}^{N}\sum_{y=1}^{M}(I_A(x,y)-I_B(x,y))^2} \qquad (33)$$

$$PSNR = 10\log_{10}(\frac{R}{RMSE})^2 \qquad (34)$$

$$AG = \frac{1}{(M-1)*(N-1)} \times$$

$$\sum_{i=1}^{N-1}\sum_{j=1}^{M-1}\sqrt{\frac{(I(i+1,j)-I(i,j))^2+(I(i,j+1)-I(i,j))^2}{2}}$$

$$(35)$$

where P_g is the probability of grey value g in the image, R is the maximum difference in the original image data type, $I_A(x,y)$ and $I_B(x,y)$ is respectively the enhanced and original gray pixel at position (x,y). MI is mean intensity value of I image. RMSE and PSNR are two error metrics used to compare the quality of original image and enhanced image. AG is average gradient reflecting the clarity of the resulted image.

Furthermore, we analyze different index data of images in Fi.1 using different methods, as data shown in Table 1. The entropy of image sing HLEPDE is greatest among five methods. Commonly, the greater the entropy

(a) Original image   (b) Image using SPLHE   (c) Image using TV

(d) Image using SPLHE before TV   (e) Image using SPLHE after TV   (f) Image using HLEPDE
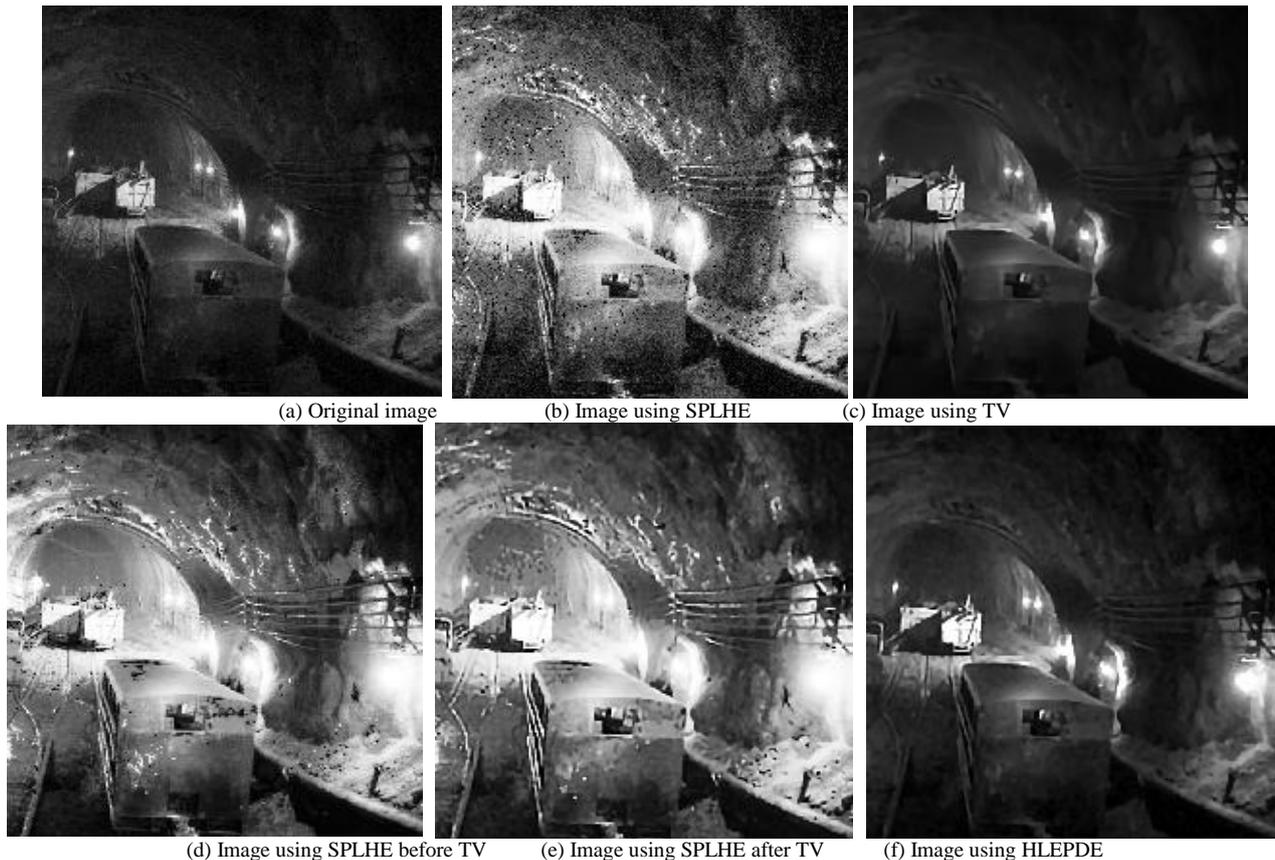
Figure 4.   Underground mine tunnel image processing using different methods

TABLE II.   INDEX DATA OF UNDERGROUND MINE TUNNEL IMAGE USING DIFFERENT METHODS

	Mean intensity	Entropy	RMSE	PSNR	Average gradient
SPLHE	107.8075	6.6002	79.6146	10.1109	20.4215
TV	41.1985	6.6444	10.9632	27.3384	2.1029
SPLHE then TV	106.7225	7.8183	79.1620	10.1605	7.1535
TV then SPLHE	105.9427	6.5563	78.4404	10.2400	5.6608
HLEPDE	48.4188	6.9060	15.1575	24.5183	10.7511

of image is, the more abundant information included in it, and the greater the quality of image is. In addition, RMSE represents the cumulative squared error between two images, whereas PSNR represents the measure of the peak error. A larger average gradient means a higher contrast. HLEPDE has worse denoising effect for Fig. 3(f) from HLEPDE has lower PSNR, but PSNR of Fig. 3(f) is greater than those of Fig. 3(b), Fig. 3(d) Fig. 3(e). Furthermore, HLEPDE can enhance higher contrast of image by comparing data in the sixth column of Table 1, but it has lower average gradient value than SPLHE. Although HLEPDE has worse contrast enhancement ability than SPLHE and worse denoising effect than TV, it has better contrast enhancement and denoising than other two methods. Therefore, HLEPDE can allow for simultaneous denoising and contrast enhancement using a trade-off scheme.

Finally, HLEPDE is applied to process image of mine tunnel (see Fig. 4(a)) with low illumination, low contrast , and some noise. Fig. 4(f) using HLEPDE has better quality than Fig. 4(a)-Fig. (e). Since illumination of Fig. 4(a) is not uniform, Fig. 4(f) cant enhance too much

illumination for keeping detailed shape, otherwise it is excessively exposed as shown in Fig. 4(b), Fig. 4(d), and Fig. (e). Fig. 4(f) has lower noise polluted and higher contrast than Fig. 4(a), but its contrast and denoising are respectively not enhanced much more than Fig. 4(b) and Fig. 4(c) in case local shape or detail is lost. Data in Table 2 also illustrates that HLEPDE enhancing image is better than other methods. In a word, HLEPDE can process image of mine tunnel for simultaneous denoising and contrast enhancement using a trade-off scheme.

IV.   CONCLUSIONS

A hybrid process method, called HLEPDE, is presented that integrated SPLHE into TV method with edge-stopping function via PDE for simultaneous denoising and contrast enhancement. The SPLHE is to enhance local contrast and preserve shape of image, while TV method with edge-stopping function is to denoise and keep edge information of image. With closing to edge of image, the smooth operation of TV is weakening. The HLEPDE is validate by processing a standard image, and further applied to a real image of mine tunnel. The results

prove that HLEPDE is effective for simultaneous denoising and contrast enhancement.

REFERENCES

[1] M. Hongwei, T. Hao, and L. Xiaopeng, "Research on Motion Control System of Mine Rescue Robot," *Proceedings of IEEE International Conference on Automation and Logistics*, Qindao, China, August 2009, pp. 1892-1895.

[2] R. Murphy, J. Kravitz, S. Stover, and R. Shoureshi, "Mobile robots in mine rescue and recovery," *IEEE Robotics & Automation Magazine*, Vol.16, pp. 91-103, February 2009.

[3] G. Tongying, D. Zhenjun, X. Fang, and W. Haichen, "Research on Localization Method of Mine Rescue Robot," *Proceedings of 2nd International Conference on Industrial and Information Systems*, Dalian, China, July 2010, pp. 503-506.

[4] Z. Rong and W. Yong, "Application of improved median filter on image processing," *Journal of Computers*, Vol.7, pp. 838-841, April 2012.

[5] K.A. Panetta, and J. Xia, "Color image enhancement based on the discrete cosine transform coefficient histogram," *Journal of Electronic Imaging*, Vol.21, pp. 1-11, February 2012.

[6] S. wenzhu, W. hongyu, and Q. Daxing, "A novel error resilient scheme for waveletbased image coding over packet networks," *Journal of Networks*, Vol.7, pp. 1046-1053, July 2012.

[7] P. Jing and M. Yan, "Integer wavelet image denoising method based on principle component analysis," *Journal of Software*, Vol.7, pp. 982-989, May 2012.

[8] Y. Zhihui, G. Fangfang, D. Ping. Robust skeleton extraction of gray images based on level set approach. Journal of Multimedia, Vol.8, No.1, pp. 24-31, 2013.

[9] Z. Shuiping, L. Huijune. The research of image encryption algorithm based on chaos cellular automata. Journal of Multimedia, Vol.7, No.1, pp. 66-73, 2012.

[10] Z. Xinming and Y. Lin, "A fast image thresholding method based on chaos optimization and recursive algorithm for two-dimensional Tsallis entropy," *Journal of Computers*, Vol. 5, pp. 1054-1061, July 2010.

[11] P. Perona, and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12, pp. 629-639, July 1990.

[12] V. Caselles, J.L. Lisani, J.M. Morel, and G. Sapiro, "Shape preserving local histogram modification," *IEEE Trans. Image Processing*, Vol. 8, pp. 220-230, January 1999.

[13] B. Begovic, V. Stankovic, and L. Stankovic, "Contrast enhancement and denoising of Poisson and Gaussian mixture noise for solar images," *Proceedings of 18th IEEE International Conference on Image Processing*, Brussels, Belguim, Septemer 2011, pp. 185-188.

[14] R.K. Jha, "Noise-induced Contrast Enhancement of Dark Images using Non-dynamic Stochastic Resonance," *Proceedings of 18th NCC, New Delhi*, India, February 2012, pp. 1-5.

[15] W. Dakai, H. Yuqing, and P. Jinye. *Partial Differential Dquations in Image Processing*, Beijing: Science Press, 2008.

Wanfeng Shang was born in 1979. She received PhD in Mechanical Engineering in 2009 from Xi'an Jiaotong University. She is currently working as a lecture in Xi'an University of Science and Technology. Her research interests are image processing and visual reorganization of robots.

Hongwei Ma was born in 1957. He received BS and MS in Mechanical Engineering in 1984 and 1993 from Xi'an University of Science and Technology, and PhD in Mechanical Engineering in 1998 from Xi'an Jiaotong University. His research interests are monitoring and control of mine device.

Xuhui Zhang was born in 1972. He received BS and MS in Mechanical Engineering in 1996 and 2002 from Xi'an University of Science and Technology, and PhD in Mechanical Engineering in 2009 from Xi'an Jiaotong University. His research interests are signal analysis and fault diagnosis of mine device.

Text Classification Retrieval Based on Complex Network and ICA Algorithm

Hongxia Li

Library of Dezhou University, Dezhou, China
Email: lihongxiadzu@yeah.net

Abstract—With the development of computer science and information technology, the library is developing toward information and network. The library digital process converts the book into digital information. The high-quality preservation and management are achieved by computer technology as well as text classification techniques. It realizes knowledge appreciation. This paper introduces complex network theory in the text classification process and put forwards the ICA semantic clustering algorithm. It realizes the independent component analysis of complex network text classification. Through the ICA clustering algorithm of independent component, it realizes character words clustering extraction of text classification. The visualization of text retrieval is improved. Finally, we make a comparative analysis of collocation algorithm and ICA clustering algorithm through text classification and keyword search experiment. The paper gives the clustering degree of algorithm and accuracy figure. Through simulation analysis, we find that ICA clustering algorithm increases by 1.2% comparing with text classification clustering degree. Accuracy can be improved by 11.1% at most. It improves the efficiency and accuracy of text classification retrieval. It also provides a theoretical reference for text retrieval classification of eBook.

Index Terms—Complex Networks, Text Classification Retrieval, ICA Clustering Algorithm, Independent Component Analysis, Collocation

I. INTRODUCTION

The modern library is developing toward information and network. The so-called information library using modern information technology to collect, organize, normative process and compress highly value images, text, voice, sound, images, video, software, scientific databases and multimedia information [1,2]. They are converted into digital information. The high-quality preservation and management are achieved by computer technology [3]. It realizes knowledge appreciation. The efficient economical transmission and reception are made by network communication technology. So that the user can get a variety of services from the net at any time and any place. Text classification is one of the key technologies for the library information of text data. Text classification has a very important significance for text information retrieval and identification. Nowadays, the most commonly used text classification is the vector space SVM model. The calculation and controllability of the model is stronger. But the method takes the word as

an independent variable. It ignores the links between words, thus affects the efficiency of text classification retrieval. With the widely use of complex network data of technology in recent years, domestic and foreign scholars have carried out research on the complex network of text classification. In which, Zhu used the length of network node average path to extract the keyword. He built text classification network, but there was deficiency in the algorithm and the extraction of keyword was not clustered [4]. Liu used statistical methods of complex networks to build a similarity network classification toward the keywords meaning of China National Knowledge Infrastructure. But the similarity of the classification is not obvious [5]. Huang used the collocations to build the complex network of the words relationship to extract keywords. Peng Zhao used the network nodes and clustering coefficient to extract keywords, but the efficiency and accuracy of the algorithm needs to be improved. These uncertain information need to rely on fuzzy sets and rough sets theory as analysis tool, especially all kinds of complicated data. Rough set and fuzzy set make up non precise information to express two important drawbacks, which are respectively indiscernible and ambiguous. The former is the essential attribute of things; the latter is a classification problem. Fuzzy set theory is proposed by Zadeh in 1965, it has been proved practical in chemistry and other disciplines. In contrast, Pawlak introduced the rough set theory in 1985, although its theory is very popular in many disciplines, it isn't mentioned in the chemical. This also shows that both the rough set theory and the other set theory have essential difference.

The traditional set theory, such as fuzzy sets, the elements in the collection can be clearly expressed. Using membership function describes common elements and collection property relations, property relations can be with or without. The definition of the membership function does not take into account the elements' uncertain problems in the collection, in order to deal with uncertainty, fuzzy sets are proposed. The fuzzy set membership function, its value can be from the closed interval of 0 to1, and allowing the segment. The fuzzy set membership function describes the events on the extent to occurs, rather than whether it occurs.

This paper introduced complex network theory in the text classification process. The classification and extraction of text classification keywords carried out by

network-weighted. The text classification aggregation was build through semantic links between words. At last, we introduced the ICA clustering algorithm. It made up for the deficiencies that the keyword can not achieve clustering [6]. The clustering divided of the keywords network node was carried out. It realized the complex network clustering classification of text classification successfully.

II. RELATED WORK

When extracted Chinese collocation words, Ruifeng Xu introduced the network-perception algorithms. Through network-perception algorithm, it will combine optimization discrimination Eigen values. Different types of collocation will be collocated at different stages respectively [7]. But this collocation relationship was not necessarily linear relationship.
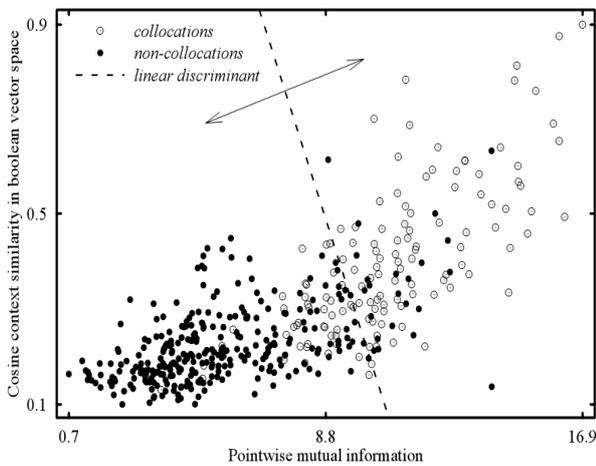

Figure 1. The linear collocation diagram of text classification search word

As is shown in Figure 1, the abscissa and ordinate represent the words collocation diagram that we get by using different statistical methods. The hollow point and solid point represent collocation and non-collocation respectively [8]. We can see from the Figure that the collocation words concentrate in the left part. So we can find the boundary of collocation and non-collocation to achieve linear collocation of the words.

The process of text classification needs to do collocation of text classification search keyword. The clustering point is used to form a complex network. Figure 2 show that the cluster has a clear dividing boundary in two-dimensional space. This means that there are two distinct clusters in two-dimensional space. If you simply follow the direction of maximum variance to project the data on longitudinal axis, you cannot separate these clustering data. So that the maximum variance cannot be true response the true semantics of text classification keywords. Therefore, it can show the authenticity of the words semantics through the clustering data rotation, and then projected onto the axis [9]. The paper put forwards ICA algorithm which is the method of rotating space. It can find the projection direction in the classified text feature word, and form the ICA semantic

space. We don't need to use the orthogonally of the two-dimensional space and reveal the semantic relationship of keywords.
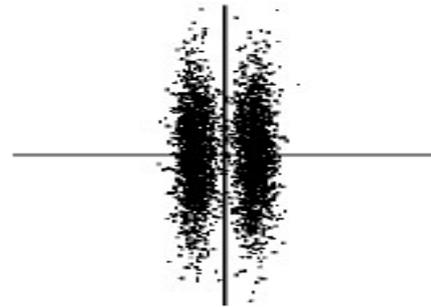
Figure 2. ICA algorithm diagram of the text classification search word

III. PROPOSED SCHEME

In the massive process of data mining, some of the data are often vague and cannot be classified. These data cannot be counted in a collection and cannot be existed in subset and the complement of a subset, but it can be counted in the boundary of the set. The study of rough set can use the similarity matrix which can be expressed as follows:

$$
\begin{matrix}
1 & a_{12} & \cdots & a_{1i} \\
a_{21} & 1 & \cdots & a_{2i} \\
\vdots & \vdots & \ddots & \vdots \\
a_{j1} & a_{j2} & \cdots & 1
\end{matrix}
\tag{1}
$$

In formula (1), a_{ji} is the value of object j according to a certain degree of similarity of the object i. If the value of a_{ji} is larger, the similarity between two objects is smaller.

The use of rough sets can be automatically controlled to extract data. This paper presents a new control strategy FRC--fuzzy-rough control. The basic idea of this control strategy is: We can use a data recording manner to record the state and representative measures in the control process and integrate these data using rough sets. We can summarize the data integration process as follows:

The rule one IF Condition 1 corresponds to THEN TAKE project 1;

The rule second IF Condition 2 corresponds to THEN TAKE project 2;

The rule third IF Condition 3 corresponds to THEN TAKE project 3.

This data processing strategy is called paradigm learning. This approach based on the coarse control and fuzzy control. Rough control has the features of simple, quick and easy. Another feature is data control algorithm comes from the data itself. Its decision-making and reasoning process are easier. It is easier than fuzzy control to inspect and operate and it is applied in a simple algorithm.

The genetic algorithm is an information data processing method based on the principle of binary data and genetic information. GA genetic test functions can be

added in the clustering algorithm for the convergence of clustering algorithm. The GA test function is shown in formula (2) [10].

$$p(y_j) = 0.02 + \sum_{i=1}^{25} \frac{1}{i + \sum_{j=1}^{2} (y_j - a_{ji})^6} \qquad (2)$$

In formula (2), a is judgment matrix. Function has more than one maximum. Generally speaking, if the function value is greater than 1 it is convergence. This test method is fast and robust performance is good.

Assuming that data samples is $b_k = (b_{k1}, b_{k2}, ...b_{kn})$, the distance between b_k and b_l can be defined as:

$$d(b_k, b_l) = \sqrt{\sum_{z=1}^{c} (b_{kz} - b_{lz})}, k, l = 1, 2, ...n \qquad (3)$$

The density of the data samples at the point of b_k can be expressed in formula (4).

$$M_k = \sum_{l=1}^{n} \frac{d(b_k, b_l)}{\sum_{k=1}^{n} d(b_l, b_e)}, k = 1, 2, ...n \qquad (4)$$

Then, we can select the next best cluster center. First, we can make the sample density distribution around more intensive. The density can be defined in formula (5).

$$Mp_k = \exp(-\frac{M_k}{\sum_{i=1}^{g} d(B_j, s_i)}), j = 1, 2, ...n, j \neq i \qquad (5)$$

The complex network is complex system that composed by structure distribution of a simple network system. The complex network of language is used to research language search structure of text classification through replication network structure. It increases the efficiency of text search [11]. Soie thought that the search statement of text classification reflected the basic characteristics of complex network. It includes the voice and lexical of text search statement, the syntax and semantics of text search statement.

A. Independent Component Analysis
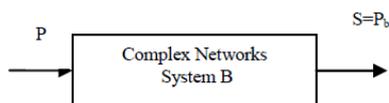
Figure 3.   The data transmission diagram of complex network system

The independent component analysis of complex network analyzes the input text classification information search source P and e-book database system B under the circumstance of no priori known. Only complex network system output data S to estimate P and B as shown in Figure 3 and Figure 4.

Assuming that P is the m-dimension vector, and then P can be expressed as $P = [P_1, P_2, \cdots, P_n]^T$.

Assuming that there is a relatively independent relationship between each component, it is shown in formula (6).

$$Q(P) = \prod_{j=1}^{n} Q(P_j) \qquad (6)$$

Text classification information search source P and complex network system B combine into n-dimension visualization vector $S = [S_1, S_2, \cdots, S_n]^T$. It can be written in the form of independent component, as is shown in formula (7).

$$S_j = \sum_{i=1}^{n} s_{ji} p_i, j = 1, 2, ..., m \qquad (7)$$

It can be written in the form of vector, as is shown in formula (8).

$$S = Pb \qquad (8)$$

ICA algorithm is based on the vector S. By solving complex network search matrix A, it can represent each search component of text classification information search source P by z. The estimation process is shown in Figure 4.
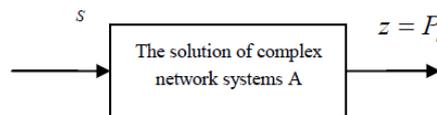
Figure 4.   The diagram of complex network solution process

In which, A is mixing matrix of m*n. z is the system output variable of search results. Generally speaking, the solution of complex network system is composed of two parts. One part is spheroidzing decomposition matrix C, the other is orthogonal decomposition matrix D. That is to say A = CD, it is shown in Figure 5.
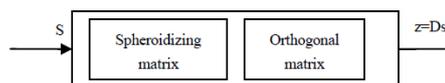
Figure 5.   The diagram of complex network matrix decomposition

B. Text Classification Algorithm

We can see from the independent component analysis of complex networks. As for text classification retrieval of complex network, we can combine with network nodes. The nodal connections are established between the keywords of different text [11]. Thereby, the keywords and information are able to reflect the related semantics and context characteristics. Through the combination of ICA classification algorithm, this paper constructs the classification algorithm steps of network text. As is shown in the follows:

Step 1: Enter the text to be classified.

Step 2: The text classification is processed. It includes sorting out the overlapping words of text classification; the removing of auxiliary verb, the removing of conjunctions words. The identification of the sentence is based on the criteria of full stop of each sentence.

Step 3: The weight of semantically related words are created. The weight can be defined in formula (9).

$$w_{tr} = \log(1 + \frac{Tre(e_i, e_r)}{Tre(e_i) + Tre(e_r) - Tre(e_i, e_r)}) \quad (9)$$

Among them, $Tre(e_i, e_i)$ represents the co-occurrence frequency of characteristic word (e_i, e_i) in the span of less than 2; $Tre(e_i)$ and $Tre(e_j)$ represent the total frequency text classification; $Tre(e_i, e_i)$ represents the high frequency of classification words of associated words. So, the higher value of $Tre(e_i)$ and $Tre(e_j)$ indicate the greater probability of classification word alone.

Step 4: The language-related search words of complex network text classification is created. We merge the characteristic words of text classification retrieval that the frequency span is less than 2.

Step 5: The weighted degree of each node is calculated. The weighting coefficient of the nodes and clustering coefficient of nodes can take weighted normalized treatment. The construction of evaluation right function is shown in formula (10).

$$W = \alpha_1 W_1 + \alpha_2 W_2 + \alpha_3 W_3 \quad (10)$$

In which, $a_i (i = 1, 2, 3)$ is adjustable parameter, it represents weight coefficient, there is $a_1 + a_2 + a_3 = 1$.

TABLE I.   A PROTEIN COMPONENT OF DECISION TABLE

U	a1	a2	a3	a4	a5	d
X0	0.23	0.54	251.2	2.215	0.02	8.3
X1	0.48	0.51	296.5	2.966	0.03	8.2
X2	0.61	1.20	287.9	2.994	1.08	8.6
X3	0.45	1.40	282.9	2.933	0.11	10.3
X4	0.11	0.29	335.0	3.458	1.19	6.5
X5	0.51	0.76	311.6	3.243	1.43	8.8
X6	0.00	0.19	315.6	2.932	1.03	7.1
X7	0.15	0.25	337.2	3.856	1.06	7.9
X8	1.20	2.10	322.6	3.350	0.04	9.9
X9	1.28	2.00	324.0	3.518	0.12	8.7

Rough set theory is applied to the composition of quantitative analysis problem, to model data source that is a protein component, wherein the amino acid of 10 coded has 5 attributes: $a_1 = PIE$ is expressed as side-chain fat-soluble; $a_2 = DGR = \Delta G$ is said that the protein convert to water content; $a_3 = SAC$ is surface area; $a_4 = MR$ is molecular refraction index; $a_5 = LAM$ is side-chain polarity.

First of all, to establish decision table, then the quantitative attributes of $\{a_1, a_2, a_3, a_4, a_5\}$ and table decision attribute $\{d\}$ constitute a decision table, which is shown in Table I.

The condition attributes are encoded as 3 quantitative intervals, such as using 1, 2 and 3 respectively expresses low, medium and high, all the attributes use natural number coding to quantitative interval, as shown in Table II.

TABLE II.   THE PROPERTIES OF QUANTITATIVE ANALYSIS

Property	Coding			
	1	2	3	4
a_1	<0.115	[0.115,0.54]	[0.54,1.195]	>1.195
a_2	<1.167	[1.167,0.75]	[0.75,1.25]	>1.25
a_3	<264.5	[264.5,321.5]	[321.5,361.2]	>361.2
a_4	<2.75	[2.75,3.54]	[3.54,4.861]	>4.861
a_5	<1.21	[1.21,1.64]	[1.64,2.15]	>2.15
d	<8.31	[8.31,12.66]	[12.66,13.98]	

All the attribute set of each molecule is upper and lower approximation, to calculate all approximate accuracy that approximately equal to 100%. Therefore, the quality of classification is also approximately 100%.

The next step of rough set analysis is to build the minimal subsets of independent attributes, to ensure the quality of classification and collection has the same effect. There are four d reductions that are given as follows:

Set #1 = $\{a_2, a_4\}$

Set #2 = $\{a_2, a_7\}$

Set #3 = $\{a_2, a_5, a_1\}$

Set #4 = $\{a_2, a_5, a_6\}$

The intersection of all d reduction is the core of the property d. In this example, the d core is $\{a_2\}$, it means the property is the most important basis for classification. In order to ensure the classification quality is not reduced, this property can't be reduction. The base of reduction can be 2 or 3, the excess is not necessary, to remove the classification quality will not be affected, minimal reduction set is #1 and #2. Therefore, the initial coding of information system attribute can be reduced from 7 to 2. This shows that based on the d core and d reduction, the relevant attributes can be further reduced. After the reduction, the information system can be viewed as a decision table. The classification accuracy of reduction set #1 is shown in Table III.

TABLE III.   THE CLASSIFICATION ACCURACY SET {A2, A4}

Category	Accuracy		
	$\{a_2, a_4\}$	$\{a_2\}$	$\{a_4\}$
1	1.000	0.361	0.162
2	1.000	0	0
3	1.000	0	0

It can be seen from the Table III, the minimum reduction set #1 = $\{a_1, a_4\}$ can ensure the classification accuracy that is 100%, while using separately a_2 or a_4 can't guarantee the classification accuracy, and it shows that this algorithm is effective and correct.

IV.   EXPERIMENTAL RESULTS

In order to verify ICA clustering algorithm of complex network text classification that proposed in this paper, we download data from the Google Web site. And the text is taken as the raw materials of classification. The text materials includes 10 different texts, they are

entertainment, publicity, advertising design, finance, sports, education, military, science and technology, legal, and news. And we take 600 articles from each text. In experiment, the 6000 articles are divided into 10 parts by the method of cross-over experiment. At first, we use the principle of independent component analysis to extract the characteristic words of the text independently. Then, the mathematical statistics of the text characteristic words of clustering degree is based on the principle of clustering algorithm. Finally, we get the statistical classification results through the key search word. The specific procedures are shown in Figure 6.

It can be seen from Figure 6, text classification process is divided into three steps:

Step1: The classification documents are sorted. The weighted complex networks of document Eigen values are constructed. The overlapping portions of characteristics words node are brought forward.

Step2: The algorithm classifier can be used to classify the classification documents. The real-time feedback of classification results are made in the classification process.

Step3: In order to achieve optimal results, we adjust the classification parameters.
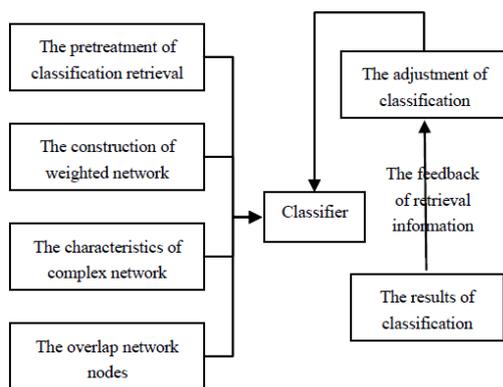
Figure 6. Txt classification process diagram of complex network

This paper choose 6000 article in the Google website to do text classification.

The characteristics of characteristic words are extracted from the results. The ICA algorithm is used to do text classification clustering. The pseudo-code of ICA algorithm is shown below:

```
get the search term Linear discriminate of a set of search term
location extracted online.
function get Search Term (initial Search Term location)
search Engineering Village using initial Search Term location, get
top 10 received results.
identify all the Search Term entities in the received results using
Lingpipe.
extract a few of Search Term of the location entities online,
```

$$F = \{F_1, F_2, \cdots, F_n\} \in K^2$$

given: a Search Term frequency

j: $K^2 \times K^2 \to K$

a function for computing the mean

l: $f(k) \to K$

select a initial centers $n_1, n_2, \cdots, n_a$

while stopping criterion is not true do

for all Term entities b_i do

$$b_i = \left\{ f_i \middle| \forall b_q j(f_j, n_i) \le j(f_j, n_i) \right\}$$

end

for all means b_i do

$n_i = l(b_i)$

end for

end

return $b_1, b_2, \cdots, b_n$

end function

As for the reliability of the algorithm, we should detect clustering degree of text classification characteristic words first. The test result of clustering quality indicator is shown in Figure 7.
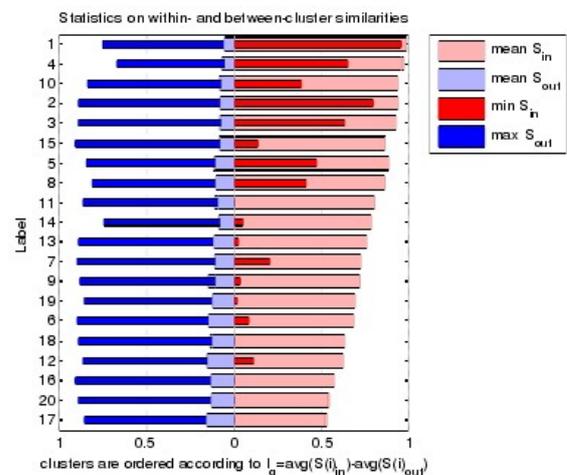
Figure 7. The results of clustering quality indicators
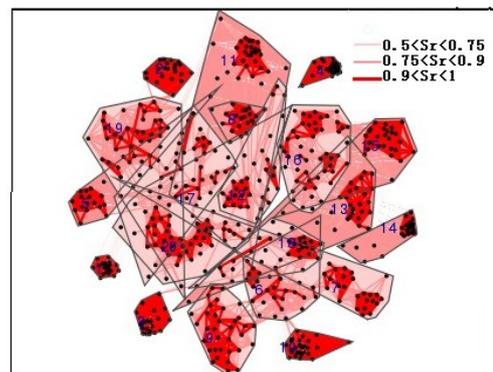
Figure 8. The statistics chart of independent component clustering of search characteristic words

In Figure 7, we can obtain the statistical results of Iq quality indicators by statistical of clustering quality indicators. In which, the compact degree of clustering can be seen from the right part of the figure. The longer of deep color part represents that the compact degree of clustering is higher. The degree of similarity between the classifications can be seen from the left part of the diagram. In which, the longer of the dark part represents that the degree of similarity between classifications is higher and the separability is better.

When the number of clustering is less, the compact degree of clustering is higher and the distinction between

classifications is better. But with the increase in the number of clustering, the clustering will become worse and worse and the similarity degree between classifications is growing. It is staggered between elements. The distinction of clustering effect is worse. It can be seen in Figure 8.
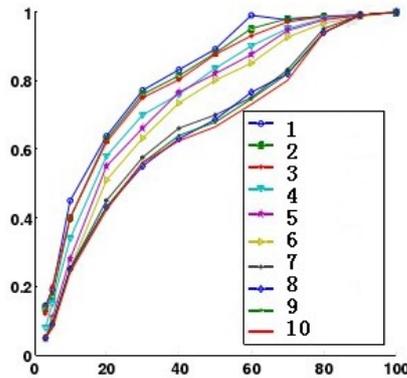
Figure 9.   Clustering degree diagram of different algorithm

Figure 8 shows the clustering degree diagram in the case of different algorithm. In order to minimize the impact between the text classifications, the paper will make cross treatment of ten texts after classification. In which, five groups will use collocation algorithm, and the other five groups will use ICA algorithm as shown in Figure 9. The clustering degree of the ten groups' text classification is quite different, in which the effect of ICA algorithm is better than collocation. In order to comprehensively verify the reliability of ICA clustering algorithm, it gives the accuracy figure of algorithm as shown in Figure 10.
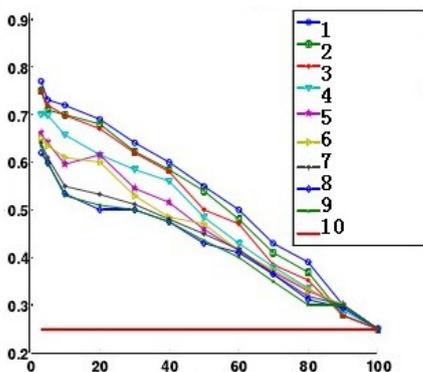
Figure 10. The accuracy figure of different algorithms

We can see from Figure 10, the different type of algorithms have different accuracy. With the increase in the number of characteristic words, the accuracy rate is significantly decreased. ICA algorithm is obviously better than collocation algorithm. In order to display the comparison results of the two algorithms in visual, the data is tabulated in the form of table as shown in Table IV.

TABLE IV.   COMPARISON RESULTS OF COLLOCATION ALGORITHM AND ICA CLUSTERING ALGORITHM

Algorithm number	Collocation clustering (%)	Collocation accuracy (%)	ICA algorithm clustering (%)	ICA algorithm accuracy (%)
1	97.2	62.1	98.2	68.1
2	98.1	63.2	99.1	70.2
3	97.3	64.4	98.3	75.3
4	98.3	65.3	99.5	76.5
5	97.2	66.1	99.3	77.2

We can see from Table IV, ICA algorithm text classification search word clustering and accuracy are obviously better than collocation algorithm. In which, the highest collocation algorithm clustering and accuracy was 98.3% and 66.1%, respectively. The highest ICA algorithm clustering and accuracy was 99.5% and 77.2%, respectively. From the results of clustering and accuracy, ICA algorithm was higher than collocation algorithm. It verified the validity and reliability of ICA algorithm programming.

V.   CONCLUSIONS

In this paper, we can combine the basic theory of complex network algorithms and conducted the depth analysis research of text classification retrieval theory. The experiment research of text classification effect is made by using the collocation and ICA algorithms. Firstly, this paper describes the principle of collocation algorithm and ICA algorithm in detail. And then it gives the construction of ICA method independent component and complex network in detail by the way of mathematical modeling. It gives the pseudo-code of the ICA algorithm. Finally, we obtain the comparison results of text classification algorithm clustering and accuracy through the algorithm experiment. In which, the highest collocation algorithm clustering and accuracy are 98.3% and 66.1%, respectively. The highest ICA algorithm clustering and accuracy are 99.5% and 77.2%, respectively. It is higher than collocation algorithm by 1.2% and 11.1%. ICA algorithm is higher than collocation algorithm. It verifies the validity and reliability of ICA algorithm programming.

REFERENCES

[1] Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler, " A review of content-based image retrieval systems in medical applications—clinical benefits and future directions", *International Journal of Medical Informatics,* vol. 73, Issue 1, pp. 1-23, February 2010.

[2] Björn Schuller, Anton Batliner, Stefan Steidl, and Dino Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge", *Speech Communication*, vol 53, Issues 9–10, pp. 1062-1087, November–December 2011.

[3] Weihong Deng, Yebin Liu, Jiani Hu, and Jun Guo, "The small sample size problem of ICA: A comparative study and analysis", *Pattern Recognition*, vol. 45, Issues 12, pp. 4438-4450, December 2012.

[4] Eduard Hovy, Roberto Navigli, and Simone Paolo Ponzetto, "Collaboratively built semi-structured content

and Artificial Intelligence: The story so far", *Signal Processing*, vol. 194, Issues 12, pp. 2-27, January 2013.

[5] Michèle Basseville, "Divergence measures for statistical data processing-An annotated bibliography", *Artificial Intelligence*, vol. 93, Issues 4, pp. 621-633, January 2013.

[6] Pedro Gil-Jiménez, Hilario Gómez-Moreno, and López-Sastre, "Evaluation of shape classification techniques based on the signature of the blob", *Signal Processing*, vol. 92, Issues 1, pp. 63-75, January 2012.

[7] Paulito Palmes, Hung Keng Pung, and Tao Gu, "Object relevance weight pattern mining for activity recognition and segmentation", *Pervasive and Mobile Computing*, vol. 6, Issues 1, pp. 43-57, February 2012.

[8] Abdullah Iqbal, Nektarios A. Valous, Da-Wen Sun, and Paul Allen, "Parsimonious classification of binary lacunarity data computed from food surface images using kernel principal component analysis and artificial neural networks", *Meat Science*, vol. 87, Issues 2, pp. 107-114, February 2011.

[9] Zhu M X, Cai Z, and Caio S, "Automatic Keywords Extraction of Chinese Document Using Small World Structure", Proceedings of the 2003 International Conference on Natural Language Processing and Knowiedge Engineering, pp. 438-443, 2010.

[10] Liu G, Zhai Z W, "Research on Keywords Extraction of Chinese Documents Based on TEXT-NET", Proceedings of the2011International Conference on Electric Information and Control Engineering, pp. 6074-6077, 2011.

[11] Huang C, Tian Y H, and Huang T J, etal, "Semantic Scoring Based on Smaii-Worid Phenomenon for Feature Seiectionin Text Mining", ADMA, pp. 636-643,2011.

Hongxia Li (1976.5-), female, Han nationality, Dezhou, Shandong Province, librarian, Working at the Library of Dezhou University, 6 professional papers have been published.

A Supervised Statistical Data Quantization Method in Machine Learning

Jing WANG

Department of Computer Science and Engineering, North China Institute of Aerospace Engineering, Langfang, China
Email: jingwang_scope@163.com

Haohan XIE, Xiangzheng ZHAO, Ye LIANG, Xihang ZHANG, and Yajun ZHANG*

School of Control and Computer Engineering, North China Electric Power University, Beijing, China
Email: xiehaohan92@163.com

Abstract—**Data quantization methods for continuous attributes play an extremely important role in artificial intelligence, data mining and machine learning because discrete values of attributes are required in most classification methods. In this paper, we present a supervised statistical data quantization method. It defines a quantization criterion based on the chi-square statistic to discover accurate merging intervals. In addition, a heuristic quantization algorithm is proposed to achieve a satisfying quantization result with the aim to improve the performance of inductive learning algorithms. Empirical experiments on UCI real data sets show that our proposed algorithm generates a better quantization scheme that improves the classification accuracy of C4.5 decision tree than existing algorithms.**

Index Terms—**Quantization, Machine Learning, Chi-Square, Naive Bayes**

I. INTRODUCTION

Quantization has played an important role in artificial intelligence, data mining, machine learning and knowledge discovery. Many induction and classification algorithms rely on discrete attributes, and need to quantify continuous attributes, i.e. to slice their domain into a finite number of intervals in order to generate attributes with a small number of distinct values. A good quantization algorithm can not only produce a concise summarization of continuous attributes to help the experts and users understand the data more easily, but also make learning more accurate and faster [1].

Existing quantization methods can be classified by different directions, among which one is to divide the quantization methods into top-down (splitting) methods and bottom-up (merging) methods. Top-down methods [2–4] start from the initial interval and recursively split it into smaller intervals, while, bottom-up methods start from the set of single value intervals and iteratively merge adjacent intervals. The Chi2-based algorithms [5–8] are well-known bottom-up quantization methods based on probability and statistics, among which the extended Chi2 algorithm (Ext-Chi2 for short in this paper) [7] is one of the most efficient algorithms for quantization of continuous attributes based on local dependence measure.

In this paper, we focus on bottom-up quantization methods and propose a statistical quantization criterion based on the analysis of chi-square related algorithms. This criterion can discover good merging criterion by considering the properties of class number and degree of freedom in chi-square statistic. In addition, a heuristic quantization algorithm is proposed to achieve a satisfying quantization result with the aim to improve the performance of inductive learning algorithms. Empirical experiments on UCI real data sets show that our proposed algorithm generates a better quantization scheme that improves the accuracy of C4.5 decision tree classification than existing algorithms.

The remainder of this paper is organized as follows. We introduce related work in Section II. Section III presents our proposed method. Experiments and performance evaluation are introduced in Section IV. Finally, we summarize our work and conclude this paper in Section V.

II. RELATED WORK

Existing quantization techniques mostly focus on processing lower dimensional datasets. Consequently, they commonly suffer from a major challenge that how to quantify high-dimensional datasets with sophisticated nonlinear structures in addition to computational problems. Liu et al. [1] present an excellent classification of current approaches in quantization along two main axes, viz., bottom-up vs. top-down, and supervised vs. unsupervised [9]. Top-down methods, such as class-attribute interdependence quantization, start from the initial interval and recursively split it into smaller intervals. Bottom-up methods, such as chi2-based quantization, begin with the set of single value intervals and iteratively merge adjacent intervals. Unsupervised methods provide no class information, such as equal-width and equal-frequency [9], KDE [10], TDE [11]. The former two techniques starting with naive methods implement with a low computational cost, and these types of methods are vulnerable to outliers and the results obtained are rather unsatisfactory in most cases. The latter two are state-of-the-art unsupervised top-down techniques, which use density estimators to select the best

cut-points and automatically adapt subintervals to the data. They determine the quantified number of intervals by the cross-validated log-likelihood. Supervised methods provide class information with each attribute value and they are much more sophisticated, such as entropy-based quantization, class-attribute interdependence methods, and the chi2-based heuristic algorithms. Entropy-based method proposed by Fayyad and Irani [12] recursively selects the cut-points on each target attribute to minimize the overall entropy and determines the appropriate number of intervals by using minimum description length principle (MDLP). Class-attribute interdependence methods are distinguished top-down supervised quantization techniques with the objective to maximize the interdependence between the class and the continuous-valued attribute and to generate a possibly minimal number of discrete intervals. The chi2-based methods are famous bottom-up supervised quantization techniques based on statistical independence. The chi-square statistic is used to determine whether the current point is to be moved or not. The most effective chi2-based algorithms, such as modified chi2 [8] and extended chi2 [7], define the measurement of distance from χ_α^2 to χ^2 as the merging criterion with a certain significance level α, where χ_α^2 is determined by selecting a desired significance level α. However, the criteria can not reflect the possibility that the intervals are merged from the probability view.

Recently, many researchers have focused on the production of new quantization methods, i.e., semi-supervised quantization (SSD) [13], refining quantization [14] and a survey [15] of quantization techniques. A. Bondu et al. [13] develop a semi-supervised method lately. It is based on the MODL framework and quantifies the numerical domain of a continuous input variable, while keeping the information relative to the prediction of classes. Refining quantization introduces a method, called »-procedure, that constructs classical partitions on the range of an attribute taking continuous values. These partitions can be seen as refinements of the ones given by the expert or the ones given by a standard method of quantization. G. Salvador et al. [15] present a survey of quantization techniques: taxonomy and empirical analysis in supervised learning. They develop a taxonomy based on the main properties pointed out in previous research, unifying the notation and including all the known methods up to date.

III. OUR QUANTIZATION METHOD

We first introduce the rough set theory [16] used in our quantization problem. Then, a novel quantization method is proposed. Finally, we give the algorithm description.

A. Rough Set Theory

Rough set theory was first proposed by Pawlak in 1980s. Now rough set is widely used in many aspects such as data mining, data analysis, knowledge discovery in database, text indexing, and so on.

In an information system, an appreciative space can be represented as the four-tuple $S = (U, A, V, F)$, where $U = \{x_1, x_2, \cdots, x_n\}$ is the universe that denotes a finite and non-empty set. $A = \{a_1, a_2, \cdots, a_m\}$ denotes a finite set of attributes. V_a denotes the domain of the attribute a. The attribute set $A = \{a_1, a_2, \cdots, a_m\}$ can be composed of decision attribute D and the condition attributes C. $V = \bigcup_{a \in A} V_a$ & $F : U \times A \to V$ is a total function such that $F(x, a) \in V_a$ for each $a \in A$, $x \in U$, called information function.

Approximation is the core concept of rough set. Let S be an information system, X a non-empty subset of U and $\varphi \neq P \subseteq A$. The P-lower approximation and the P-upper approximation of X in S are defined, respectively, by:

Definition 1 If two elements x, $y \in U$ and $P \subseteq A$, θ_P is equivalent relation on U, if $x \theta_P y \Leftrightarrow (\forall p \in P)(f_p(x) = f_p(y))$, we say that θ_P is indistinguishable.

Definition 2 Set $X \subseteq U$, $P \subseteq C$, P-lower approximation about X can represent:

$$P_{-}X = \{x \in U \mid [x]_P \subseteq X\} \tag{1}$$

where $[x]_P$ expresses equal kind of element set under the equivalent relation P.

Definition 3 Set $X \subseteq U$, $P \subseteq C$, P-upper approximation about X can represent:

$$P^{-}X = \{x \in U \mid [x]_P \cap X \neq \varphi\} \tag{2}$$

Definition 4 Set U denote the universe, P and Q are two equivalent relations bunches on U,
Let us define $POS_P(Q)$ as follows:

$$POS_P(Q) = \bigcup_{X \in U/Q} P_{-}(X) \tag{3}$$

Definition 5 Set $P \subseteq C$, P-approximate precision (level of consistency) under partition $\{Y_1, Y_2, \cdots, Y_k\}$ is defined:

$$Lc_P = \sum_{i=1}^{k} card(P_{-}Y_i) / card(U) \tag{4}$$

where, $card()$ expresses the cardinal number of set and Lc_P reflects the correct degree of decision classification.

B. Our Proposed Quantization Method

In this section, we propose a supervised statistical chi-square algorithm to discover quantization scheme. Before getting into the details of our approach, we first state the problem of quantization. A quantization task requires a training data consisting of N samples, where each sample belongs to only one of S classes. Next, there exists a quantization scheme P, which quantifies the continuous domain of attribute into I intervals bounded by the pairs of numbers:

$$P : \left\{ [t_0, t_1], [t_1, t_2], \cdots, [t_{I-1}, t_I] \right\}$$

where t_0 is the minimal value and t_I is the maximal value of a continuous attribute. The values in P are arranged in ascending order. For the purpose of quantization, the entire dataset is projected onto the targeted continuous attribute. The result of such a projection is a two dimensional *contingency table*, see TABLE I, with I rows and S columns. Each row corresponds to an initial data interval, and each column corresponds to a different class. N_{ij} represents the number of samples with jth class in the ith interval T_i. $N_{.j}$ is the total number of samples belonging to the jth class. $N_{i.}$ is the total number of samples that are within the interval T_i.

TABLE I.   NOTATIONS OF CONTINGENCY

Intervals	Class label				Sum of Row
	Class 1	Class 2	…	Class S	
$T1 : [t0, t1]$	N_{11}	N_{12}	…	N_{1S}	$N_{1.}$
$T2 : [t1, t2]$	N_{21}	N_{22}	…	N_{2S}	$N_{2.}$
…	…	…	…	…	…
$TI : [tI\text{-}1, tI]$	N_{I1}	N_{I2}	…	N_{IS}	$N_{I.}$
Sum of Column	$N_{.1}$	$N_{.2}$	…	$N_{.S}$	$N(total)$

As we know, the chi2-based algorithms mentioned in Section II are well-known bottom-up quantization techniques based on statistical independence, among which the modified chi2 and extended chi2 are two algorithms of the most effective techniques for quantization of continuous attributes. The modified chi2 uses the measurement of distance from χ_α^2 to χ^2 as the merging criterion with a certain significance level α to determine whether the current interval pair is to be merged or not. The merging criteria can be written as follows:

$$L = \chi_\alpha^2 - \chi^2 \qquad (5)$$

where χ_α^2 is determined by selecting a desired significance level α. The value of χ^2 can be interpreted as a distance to the hypothesis of independence between the target class and the quantified attribute, which can be calculated as follows [8]:

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{s} \frac{\left(N_{ij} - E_{ij}\right)^2}{N_{ij}} \qquad (6)$$

where $E_{ij} = \dfrac{N_{i.} \times N_{.j}}{N}$ is the expected frequency of N_{ij}, and $N = \sum_{i=1}^{I} N_{i.}$. An visual presentation for chi-square distribution can be found in Figure 1.

In (2), $\dfrac{N_{.j}}{N}$ is the proportion of number of samples in jth class accounting for total number of samples, E_{ij} is number of patterns in the ith interval, jth class in such a proportion (probability). Therefore, statistic $\hat{A}2$ indicates

the equality degree of the jth class distribution of adjacent two intervals. The smaller χ^2 value is, the more similar is class distribution, and the more unimportant cut point is. So, it should be merged.
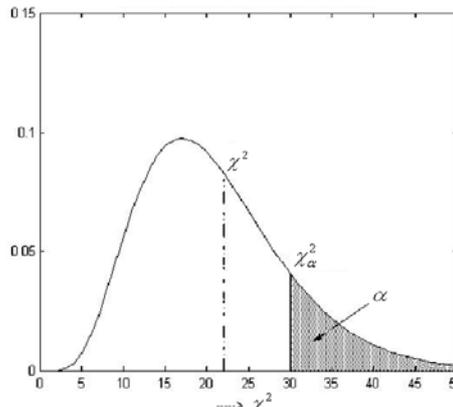
Figure 1. χ^2 distribution function

For statistic χ^2, the smaller χ^2 value is, the more similar is class distribution, the more unimportant is node. Determining the importance of a cut point by $L = \chi_\alpha^2 - \chi^2$ intends to farthest enhance statistic of nodes in system under definite significance level. χ^2 value not only relates with the number of degrees of freedom of adjacent two intervals, but also relates with the number of system classes, namely the higher the number of system classes are, the more number of classes of adjacent two intervals are, the greater degrees of freedom v of adjacent two intervals are. Thus, χ^2 value calculated is different. So, the significant quantization criterion should be determined as follows:

$$D = \chi_\alpha^2 - \frac{v}{S} \chi^2 \qquad (7)$$

Actually, this important criterion readjusts the order of interval merger after multiplying $\dfrac{v}{S}$ in front of χ^2 and reduces χ^2 value pro rata, not decreasing prematurely significance level α and illuminating adequately rationality of node merging. Thus, the new algorithm regards a new merging criterion $D = \chi_\alpha^2 - \dfrac{v}{S} \chi^2$ as basis of interval merger.

C. Algorithm

Based on the above proposition, we design an algorithm that is heuristic in nature. The objective of this algorithm is to find better quantization scheme including greedy bottom-up quantization. The detailed steps can be described as shown in Algorithm 1 of TABLE II.

TABLE II.   FLOW OF THE ALGORITHM

Algorithm 1 Our Proposed Quantization Algorithm
1: **Input:** data set with N examples and S target classes
2: **Output:** quantization scheme with a set of intervals for each continuous attribute
3: Initialize:
4: Set the significance level as α = 0.5 and Calculate the level of consistency of the original data
5: Sort attribute values in ascending order
6: for *each attribute* do
7:      Sort attribute values in ascending order
8:      Compute χ^2 value and corresponding threshold of each two adjacent intervals
9: end for
10: D(data);
11: Phase 1: Merge intervals considering all the attributes
12: while α *can be decreased* do
13:     while *Merge*(*data*) do
14:         if the level of consistency is changed then
15:             *SubD*(att);
16:         else
17:             Goto line 22
18:         end if
19:     end while
20: $\alpha_0 = \alpha$; decreasing the significance level ® by one level
21: end while
22: Phase 2: Refine intervals considering single attribute according to parallel quantization strategy:
23: Set $SL[i] = \alpha_0$
24: Do until no attribute can be merged
25: for *each mergeable attribute* do
26:     while $SL[i]$ *can be decreased* do
27:         while *Merge*(*att*) do
28:             if the level of consistency is changed then
29:                 *SubD*(att);
30:             end if
31:         end while
32:         Decreasing the significance level $SL[i]$ by one level
33:     end while
34: end for

where

Merge(*data*): considering each interval pair for all the attributes in the subspace. Merge the adjacent two intervals which have the maximal D value. Returning true or false depending on whether the concerned interval pair satisfies this condition or not;

D(*data*): calculation of D of all the interval pairs;

SubD(*att*): calculation of D which needs update;

Merge(*att*): the same as *Merge*(*data*) except for considering each interval pair for a single attribute.

The computational complexity of *Algorithm 1* is analyzed as follows. Each sort needs $O(N\log N)$. In addition, the complexity of searching best merge is $O(N^2)$, and the computation of the level of consistency also needs $O(N^2)$. Therefore, the computational complexity of *Algorithm 1* is $O(N^2)$.

IV.   EXPERIMENTAL RESULTS

In order to evaluate our proposed algorithm in a real-world situation, 11 data sets are selected from the UC Irvine machine learning data repository [17] with numeric features and varying data sizes. The data are fully consistency or correct (inconsistency rate is zero), and the data contain real-life information from the medical and scientific fields which had been used previously in testing pattern recognition and machine learning methods. A summary of data sets can be found in TABLE III.

TABLE III.   THE SUMMARY OF DATA SETS

Data sets	Number of Continuous attributes	Number of Discrete attributes	Number of Classes	samples
CPS	4	6	2	534
Iris	4	0	3	150
Auto	5	2	3	392
Breast	9	0	2	683
Ionosphere	32	2	2	351
Pima	8	0	2	768
Glass	9	0	6	214
Wine	13	0	3	178
Machine	7	0	8	209
Heart	6	7	2	296
Sonar	60	0	2	208

We compare our proposed method with the following algorithms for performance evaluation.

Built-in: C4.5 built-in quantization method [17];

Mod-chi2 and Ext-Chi2: two popular bottom-up algorithms [7, 8];

CAIM: a novel top-down method [3];

Boolean: a boolean reasoning approach [19];

MDLP: entropy-based method using the minimum description length principle [12];

Equal-F: a typical unsupervised top-down method [20].

Among the eight quantization algorithms, Equal-F, Mod-Chi2, Ext-Chi2 and our method require the user to specify in advance some parameters of quantization. For Mod-Chi2, Ext-Chi2 and our method, we set the level of significance to 0:9999 in our method. For Equal-F, the number of intervals is set to 10. MDLP has an automatic stopping rule and does not require any parameter setting.

In the following experiments, each data set is quantified respectively by the eight algorithms mentioned above. The 5-fold cross-validation test method is applied to all data sets. Each data set is divided into five parts, among which four parts are used as the training sets and one as the testing set. The experiments are repeated many times. The final predictive accuracy is taken as the average predictive accuracy value.

The learning accuracy of these eight algorithms are presented in TABLE IV. The comparison results in Table IV show that on the average, our method achieves the highest classification accuracy, which demonstrates that our method can produce a high quality quantization scheme. We rank the algorithms for each data set separately, the algorithm with the best performance gets the rank of 1; the second best gets the rank of 2, and so on. Quick comparisons of the eight algorithms can be obtained by checking the mean rank in Figure 2. Our method achieves highest mean rank.

We also select 8 UCI data sets to test C4.5 number of rules for different quantization algorithms. As seen from Figure 3 and 4, compared our algorithm with Mod-Chi2 algorithm and Ext-Chi2 algorithm, the average number of nodes of decision tree of our algorithm and the average number of rules extracted of our algorithm have decreased apparently except for Pima data set, which the two aspects are exactly the results we expect, manifesting advantage of our algorithm.

TABLE IV.   THE LEARNING ACCURACY (PERCENT) OF C4.5 DECISION TREE CLASSIFICATION

Data sets	Quantization algorithms							
	Built-in	Mod-Chi2	Ext-Chi2	Our method	CAIM	Boolean	MDLP	Equal-F
CPS	59.0	61.6	64.96	64.96	**65.85**	63.65	61.64	56.14
Iris	94.4	**94.67**	**94.67**	**94.67**	92.67	90.0	93.33	92.67
Auto	77.84	80.65	81.42	**85.64**	75.83	77.06	79.89	72.55
Breast	94.28	96.37	96.37	**97.4**	96.77	95.29	96.03	93.83
Ionosphere	90.01	**92.93**	**92.93**	**92.93**	92.01	89.72	86.87	82.6
Pima	71.69	72.95	72.95	**76.23**	73.41	66.82	70.31	66.24
Glass	34.06	43.81	43.26	**49.12**	44.29	40.48	42.15	39.87
Wine	81.53	93.53	96.47	**97.1**	92.35	88.24	83.32	90.75
Machine	79.9	88.24	89.24	**90.08**	81.86	84.43	80.4	72.0
Heart	70.41	73.23	73.23	**77.48**	74.79	69.96	74.84	69.67
Sonar	56.43	59.57	59.57	**61.03**	60.07	56.79	59.93	51.64
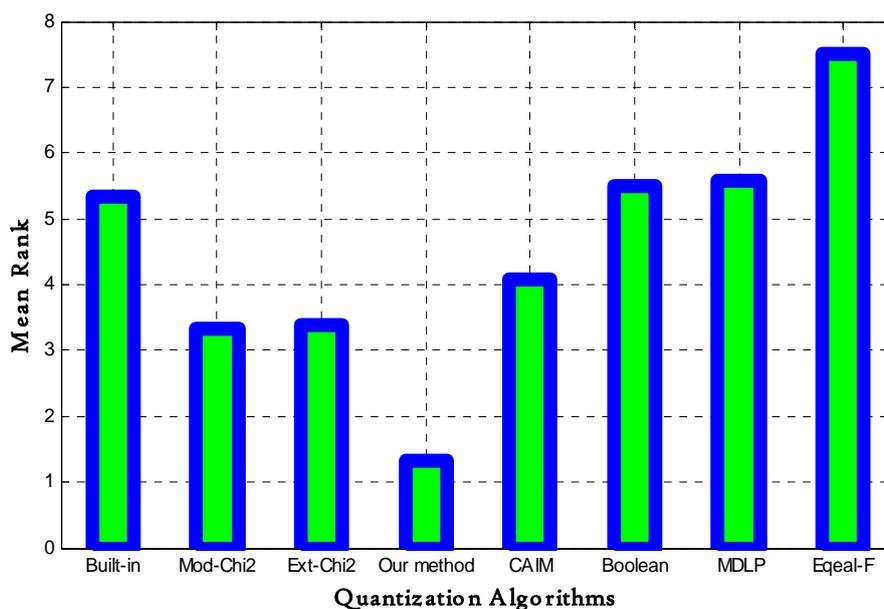
Figure 2.   Comparison of mean rank for different quantization algorithms
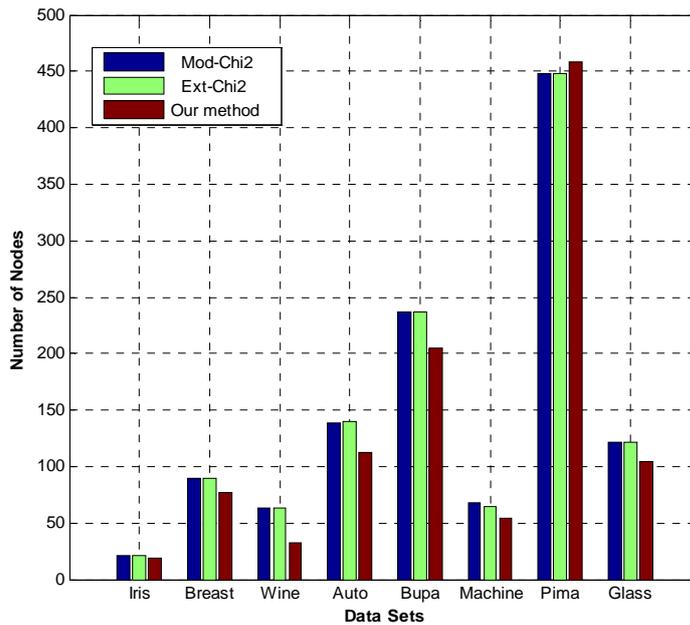
Figure 3.   Comparison of C4.5 number of nodes for different quantization algorithms
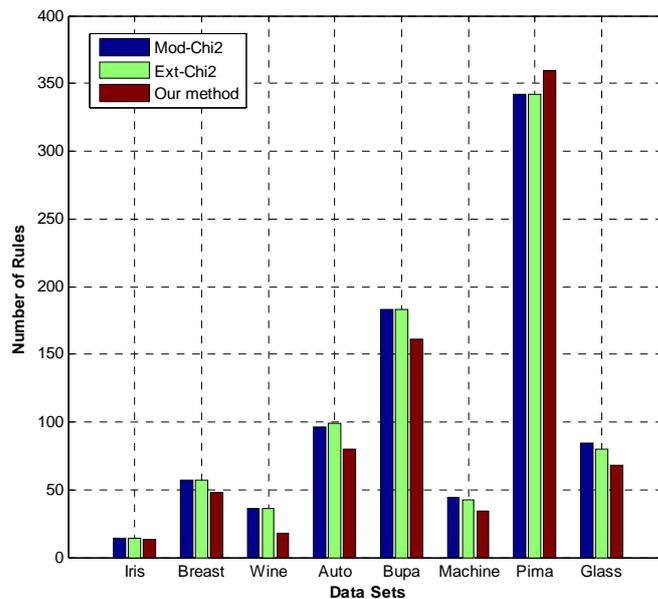
Figure 4.   Comparison of C4.5 number of rules for different quantization algorithms

V.   CONCLUSIONS

We proposed a static, incremental, supervised and bottom-up quantization algorithm in this paper, which presents a new quantization criterion and a heuristic algorithm. In order to estimate the effect of generated quantization schemes on the performance of the classification algorithm, empirical evaluation of eight quantization algorithms on UCI real data sets shows that our proposed generates a better quantization scheme to obtain the improvement of accuracy of C4.5 decision tree.

REFERENCES

[1] H. Liu, F. Hussain, C. L. Tan, and M. Dash, "Discretization: an enabling technique," *Journal of Data Mining and Knowledge Discovery*, vol. 6, no. 4, pp. 393–423, 2002.

[2]   J. Y. Ching, A. K. C. Wong, and K. C. C. Chan, "Class-dependent discretization for inductive learning from continuous and mixed-mode data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 7, pp. 641–651, 1995.

[3]   L. A. Kurgan and K. J. Cios, "CAIM discretization algorithm," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 2, pp. 145–153, 2004.

[4]   J. Catlett, "On changing continuous attributes into ordered discrete attributes," *In Proceedings of Fifth European Working Session on Learning. Berlin: Springer-Verlag*, pp. 164–177, 1991.

[5]   R. Kerber, "Chimerge: Discretization of numeric attributes," *Proceedings Ninth National Conference on Artificial Intelligence, AAAI Press*, pp. 123–128, 1992.

[6]   H. Liu and R. Setiono, "Feature selection via discretization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 9, no. 4, pp. 642–645, 1997.

[7]   C. T. Su and J. H. Hsu, "An extended chi2 algorithm for discretization of real value attributes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 3, pp. 437–441, 2005.

[8]   E. H. Tay and L. Shen, "A modified chi2 algorithm for discretization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 3, pp. 666–670, 2002.

[9]   J. Dougherty, R. Kohavi, and M. Sahami, "Supervised and unsupervised discretization of continuous feature," *Machine learning: Proc. 12th Intl Conf.*, pp. 194–202, 1995.

[10] M. Biba, F. Esposito, S. Ferilli, N. D. Mauro, and T. Basile, "Unsupervised discretization using kernel density estimation," *the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 696–701, 2007.

[11] G. Schmidberger and E. Frank, "Unsupervised discretization using tree-based density estimation," *The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, pp. 240–251, 2005.

[12] U. Fayyad and K. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," *In Proc. Thirteenth International Joint Conference on Artificial Intelligence. San Mateo, CA: Morgan Kaufmann*, pp. 1022–1027, 1993.

[13] A. Bondu, M. Boulle, V. Lemaire, S. Loiseau, and B. Duval, "A non-parametric semi-supervised discretization method," *2008 Eighth IEEE International Conference on Data Mining (ICDM)*, pp. 53–62, 2008.

[14] E. Armengol and A. Garcia-Cerdana, "Refining discretizations of continuous-valued attributes," *Modeling of Decisions of Artificial Intelligence Conference, LNAI, Springer, Heidelberg*, pp. 258–269, 2012.

[15] G. Salvador, L. Julian, S. J. Antonio, L. Victoria, and H. Francisco, "A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 734–750, 2013.

[16] Z. Pawlak, Rough set. International Journal of Information and Computer Science. vol. 11, no. 5, pp. 341–356, 1982.

[17] S. Hettich and S. D. Bay, "The UCI KDD archive [db/ol]," *http://kdd.ics.uci.edu/*, 1999.

[18] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, Calif.: Morgan Kaufmann, 1993.

[19] H. S. Nguyen and A. Skowron, "Quantization of real values attributes, rough set and boolean reasoning approaches," *In: Proceedings of the 2nd Joint Annual Conference on Information Science, Wrightsville Beach, NC*, pp. 34–37, 1995.

[20] D. Chiu, A. Wong, and B. Cheung, "Information discovery through hierarchical maximum entropy discretization and synthesis," *in: G. Piatetsky-Shapiro, W. J. Frawley (Eds.), Knowledge Discovery in Databases, AAAI Press*, pp. 125–140, 1991.

Similarity Matching Method for Music Melody Retrieval

Wang Yunjing
Capital Normal University, Beijing, China
Email: yunjingwang@163.com

Abstract—**In the research on content-based music retrieval, the fixed length of window is generally used to extract the features of melody. It has separated one fragment of music melody into multiple shorter window data. This will lose the whole feature of music melody and cause lower accuracy of retrieval. Therefore, current content-based music retrieval algorithms are discussed in detail and the emphasis is focused on the variable-length segmentation algorithm. The music retrieval algorithms for music fragments and humming fragments based on it are also analyzed. The former obtains the melody features of music fragments through accurate tonal values, calculated by the pitch extraction of single instrumental signals. The latter gets the pitch period of the humming fragments by calculation, which is based on the variable-length segmentation algorithm. Then the pitch period of each tone calculated by music retrieval algorithm and the results in experiments are used for similarity calculation. Compared to existing studies, the improved algorithm not only maintains the feature of low hit rate of variable-length segmentation algorithm, it also improves the matching speed and acquires better ability in fault-tolerance.**

Index Terms—**Music Retrieval, Variable-Length Segmentation Algorithm, Humming Melody, Similarity, Feature Matrix**

I. INTRODUCTION

With the increase of music resource on internet and the demand of people for music retrieval, the limitation of music retrieval based on text mark is more obvious. Thus, more research are focused on the retrieval based on content. Different from traditional retrieval methods, the retrieval based on content analyzes and processes the music content to extract the features related to the music content. The features are used for retrieval matching to ensure the results to meet the users' demand. Traditional methods use the retrieval technology based on text, that is the information of key words like file name, author, lyrics etc. Simply using the text to describe the audio information causes the limitations unable to overcome, they are expressed in the following aspects:

The amount of audio data is large gets becoming explosive growth. The text annotation in early work has big workload and it spends more cost. So such large scale of audio data can not be marked completely.

The audio data is a kind of binary flow, which lacks description of semantics. The mark based on text has shortcoming in subjectivity and incompleteness. So the perception of the audio is difficult to express clearly.

We can not retrieve the same file with specified audio data.

The users will not keep a brand-new memory to the audio key word all the time, Then the retrieva based on text can not meet their demand.

The audio retrieval technologies based on content is proposed to solve above problems. They extract the semantic and feature of the objects in the audio database. Then according to these information, a large amount data stored in the databased are searched to find out the audio data with similar features. Ghias and R. J. McNab etc. [1-3] carried out their research on music retrieval early. They used time domain autocorrelation method to extract pitches. According to the change of pitches the melodies are described with symbol strings (U, D, S or U, D, R). The tempo features are neglected and matching engine adopts the approximate string comparison algorithm. such as DP method or Fast approximate matching method [4, 5]. Due to the computation ability and the limitation of music library of the time, they did not give convincing test data of performance in the references. R. J. McNab [6] of Waikato university studied the retrieval technology on large-scale music library based on the work of Ghias. He uses Gold-Rabiner algorithm to detect the pitch of humming signals and adopts dynamic programming to match strings. But with the expandition of database, it is difficult to determine a song with this method. Rho [7], Marolt [8], Chang [9] and Parker [10] improved the method of R. J. McNab and they proposed based on trees. It can reduce the computing times for matching, which also improves the accuracy and speed of system. The most widely used approximate string matching algorithms [11-14] have obvious defects in fault tolerance, so most of the humming retrieval system adopt the magnitude of pitch change to describe the melody of songs. DTW algorithm [15] directly calculates time series of pitch match, it has better performance in fuzzy matching. But it adopts the dynamic programming technology, which means bigger computation and it weakens the response ability of system. The algorithm based on geometric similarity [16] extracts the pitch curves by detecting input audios. Then it compares the geometric similarity of two pitch curves in two-dimensional space. But it also needs bigger computation so it can not widely use.

In the music retrieval based on content, the users can use all kinds of methods for retrieval. Here the research focuses on the retrieval needed music files in the database through music fragment or humming fragment. The music fragment includes origin music fragment and the fragment played by other instruments which has the same melody. Since different people will get different humming tones facing the same melody and the change of tones when humming can not be mastered correctly. So the retrieval based on humming has difference with that based on music fragments. We have improved the variable-length segmentation algorithm. According to the feature matrix described in this algorithm to extract melody feature. Then new feature are formed for retrieval by similarity matching. The music retrieval based on humming calculates the similarity by fuzzy similarity matching algorithm to make music retrieval. The former can get accurate tone by the calculation of pitch extraction of single instrumental signal. Then it adopts exact similarity matching calculation method to calculate the similarity. The latter gets the pitch period of all tones of the humming fragments by calculation. It performs the fuzzy matching similarity calculation according to the feature matrix's structure extracted by variable-length segmentation algorithm. At last our experiments show the proposed method is effective and feasible

II. VARIABLE-LENGTH SEGMENTATION ALGORITHM BASED ON FRAGMENTS

A. Retrieval based on Music Fragments

The basic functions of music retrieval based on content are based on music fragments with the same melodies played by other music instruments. That is, the retrieval is processed through music content. Meanwhile, it is also the foundation of music retrieval algorithm based on humming. Since the feature matrix information extracted by whole song fragments is stored in music database, the fragments melody played by piano can be used to retrieve music files. Music retrieval with music fragments mainly includes following steps [17]. First, the database needs to store all feature matrix information of required retrieved music files. Second, signal analysis of input music fragments is performed and its tone information is extracted, so further extraction of its corresponding melody features is used for feature matrix. Finally, retrieval results by similarity matching are returned. Here we need to perform detailed description of the algorithm and process in the second step first. That is, we carry out signal analysis of input music fragments to extract its melody feature.

1) Preprocessing of music fragments

The fragment signal of playing piano is a mono WAV file with 8-bit per byte whose sampling rate is 11. 025Hz. During the filter of music fragment, in order to eliminate current noise of piano fragments, the digital low-pass filter of Butteworth is used to filter music fragments. The filter order is 10 and the cutoff frequency is $006r/s$, as is shown by following figures. Figure 1 refers to the

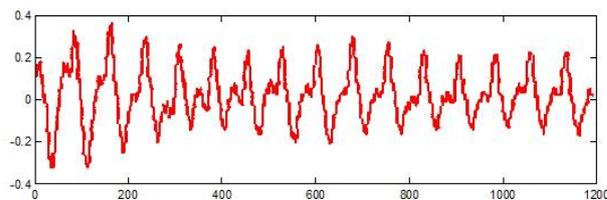signal of piano fragments and figure 2 refers to the signal after filtering.

Figure 1.   The signal of piano playing fragment.
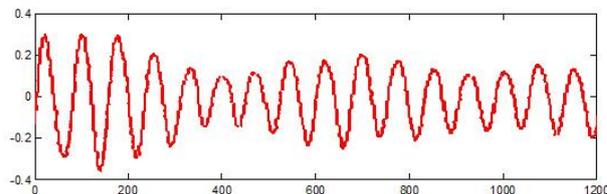
Figure 2.   The filtered signal.

During the windowed packetization of music fragments, the hamming window is used for windowed packetization. As is verified in the experiments, the choice of length of hamming window will affect the accuracy of pitch detection. When the window length is 256, the effect of pitch detection is the best. Therefore, during systematic realization, the hamming window whose length is 256 is applied to make windowed packetization of piano fragments after filtering. The rafting length between frames is 128.

2) Tone extraction of music fragments

Step 1: The signal of the first window is centric fragmentped.

Step 2: Calculate the autocorrelation function of the centric fragmentping signals and extract the peak value. Then the distance between two peaks is calculated.

Step 3: The pitch period of all windows are calculated based on this method to form the curve of pitch period value of the whole music fragments.

Step 4: Smoothing process. In this process, five-dot middle value smoothing algorithm is adopted.

Step 5: Calculate the starting point and ending point of each tone on music fragments signals and determine the number of tone music fragments.

Step 6: Calculate average pitch period value of each tone, according to starting point and ending point of the calculated pitch period curve.

3) Melody feature extraction of music fragments

The feature matrix extraction algorithm described in the variable-length segmentation algorithm [18] can be used to extract melodic features of the piano fragment signals, to establish the feature matrix. Then the music file is retrieved by similarity matching.

B. Music Retrieval based on Humming

Music retrieval algorithm based on humming is similar to music retrieval algorithm based on piano playing fragments. However, due to inaccurate tone of the humming signal, which causes some algorithms are not the same. The music retrieval algorithm based on

humming also contains two parts: signal preprocessing and tone extraction.

1) Preprocessing of humming signal fragments

We use microphone to record the humming signals in the experiment and store its signals as mono WAV files with $8bit/byte$ and the sampling rate is 11. 025 Hz. During filtering of humming signal fragments, in order to eliminate current noise, digital low-passing filter of Buttewrohrt is also applied to filter music fragments. Its filter order is 10 and the cutoff frequency is $0.06r/s$. Since the average power spectrum of humming signal is affected by glottal pulses and oral nasal radiation. The high frequency end will get voltage sag based at $6dB/fm$ above $800Hz$, so the signals will be pre-emphasised.

During the windowed packetization of humming fragments, hamming window is also used for windowed packetization. The experiments show when window length is 128, the pitch detection effect of humming signals is the best. Therefore, during systematic realization, the hamming window with 128 window length is applied to perform windowed packetization of humming fragment signal after filtering and enhancing. The rafting length between frames is 64.

2) Tone extraction of humming fragments

Since each person's humming tone is not accurate. Simultaneously, the pitch period of the same tone between people on humming cannot determine its division. So the average pitch period of each calculated tone on humming fragments cannot be transformed into determined pitch value, according to corresponding scope like piano fragments processing. We use microphone to record the humming signals in the experiment and store its signals as mono WAV files with $8bit/byte$ and the sampling rate is 11. 025 Hz. During filtering of the humming signal fragments, in order to eliminate current noise, digital low-pass fitler of Buttewrohrt is also applied to filter music fragments. Its filter order is 10 and its cutoff frequency is $0.06r/s$. Similarly, the experiment shows that the scope of pitch period of each tone is not evenly distributed.

The signal feature of humming fragments cannot determine its correct melody feature matrix. Thus, the calculated pitch period of each tone can only be applied with discovery in the experiment. Its similarity can be calculated through fuzzy matching similarity algorithm of humming fragments so the music retrieval based on humming fragments can be acquired.

III. SIMILARITY MATCHING ALGORITHM

A. Similarity Computation

For the quantized features of melody outline, such as length and tone, the similarity computation will be applied to compare input melody features and the redord in database. Since the similarity computation is rather large, the matching retrieval of strings is usually processed at first and the hitting results will be performed similarity computation again [19].

Suppose $\{b_i\}$ and $\{d_i\}$ are respectively the length feature sequence of input melody and hitting result. $\{u\}$ and $\{v\}$ are intensity feature sequence of them, $i=1,2,...,N$. N refers to the length of feature sequence. The following formula can be used to calculate intensity and similarity of length .

$$rank_d = \prod_{i=0}^{N-1}(1-\alpha\frac{|b_i-d_i|}{\max(b_i,d_i)}) \qquad (1)$$

$$rank_v = \frac{\sum_{i=0}^{N-1}(u_i-\overline{u})*(v_i-\overline{v})}{\sqrt{\sum_{i=0}^{N-1}(u_i-\overline{u})^2}\sqrt{\sum_{i=0}^{N-1}(v_i-\overline{v})^2}} \qquad (2)$$

$rank_d$ and $rank_v$ individually refers to similarity of length and intensity. α is the weight value, u and v refers to the mean value of $\{b_i\}$ and $\{d_i\}$.

$$\overline{u}=\frac{1}{N}\sum_{i=0}^{N-1}u_i, \quad \overline{v}=\frac{1}{N}\sum_{i=0}^{N-1}v_i \qquad (3)$$

With the results by similarity computation, the correlation of hitting result can be acquired by formula 4:

$$rank=(\beta\cdot rank_d)+(1-\beta)\cdot rank_v\cdot\frac{N_t}{N} \qquad (4)$$

N_t refers to the times of input, elimination and replacing operation when matching retrieval of pitch. β is the weight which determines the relative importance of length feature and the intensity feature on melody. For example, during humming, the intensity feature hold by common person is worse than his length feature. Thus, during correlation computation, the weight of intensity should be smaller than the length feature. Final determination of weight needs amounts of test and regulation. Finally, the hitting results will be sorted according to computed correlation and the results will be returned to users.

B. Similarity Matching of Music Fragment Retrieval

Similarity matching is the final step in music retrieval based on music fragments.

First, all the melody feature information of music fragments is stored in feature database. Next, calculate the feature matrix of music fragments through music retrieval algorithm of music fragments. Then, similarity matching algorithm is needed and retrieval results are returned. Since retrieval algorithm of piano fragments calculated can identify the melody value and melody feature matrix, the similarity computation of accurate match is adopted for similarity matching.

Similarity computation of exact matching on piano fragments is to perform similarity computation of exact matching music file to get similarity value. Then, sorted as similarity, returned highest music file information is taken as retrieval results. However, the algorithm of exact

matching similarity computation for each music file is shown as below:

Step 1: Extract the feature matrix of the first window in music file. If the window length is similar to that of piano fragments, follow the procedures below:

Set $W[i][j]$ as windowed feature matrix and $F[i][j]$ refers to feature matrix of piano fragments with $0 < i \leq 3$, $0 < j \leq 20$. $N = W[l][l]$ refers to window length.

α refers to adjustable parameter and $\alpha = 0.05$. S denotes *similarity* and *similarity* refers to similarity value. If window length is larger than that of piano fragment, it indicates that they don't completely match. $similarity = 0$. If windowed length is smaller than that of piano fragment, it means that there are two or more than two piano fragments without instrument intervals like the first sentence in the song "honey". According to rhythm division, it will be divided into two windows. However, people are used to taking them as one sentence. Therefore, it needs to calculate as following steps.

Step 1: Extract feature matrix of next window W_2. $N_1 = W_1[l][l]$ refers to the first window length, $N_2 = W_2[l][l]$ refers to the second window length and $N_f = F[l][l]$ refers to the length of piano fragments. If $N_1 + N_2 > N_f$, it means that they are not matching completely. So *similarity* is 0. If $N_1 + N_2 = N_f$, it needs the following calculation.

$$S = \alpha + \frac{\sum_{i=2}^{4} |W[1][i] - V[1][i]|}{2N} + \frac{2\sum_{i=2}^{N} |20 - W[2][i] - V[2][i]|}{100(N-1)}$$
$$+ \frac{\sum_{i=2}^{N} |20 - W[3][i] - V[3][i]|}{100(N-1)} \tag{5}$$

$$S = \alpha + \frac{\max(W_1[1][2], W_2[1][2] - F[1][2])}{2(N_1 + N_2)} + \frac{\sum_{i=3}^{4} |W_1[1][i], W_2[1][i] - F[1][i]|}{2(N_1 + N_2)}$$
$$+ \left(\frac{2\sum_{i=2}^{N_1} |20 - W_1[2][i] - F[2][i]| + 2\sum_{i=2}^{N_2} |20 - W_2[2][i] - F[2][i + N_1]|}{100(N_1 + N_2 - 2)} \right) \tag{6}$$
$$+ \left(\frac{2\sum_{i=2}^{N_1} |20 - W_1[3][i] - F[3][i]| + 2\sum_{i=2}^{N_2} |20 - W_2[3][l] - W_1[3][l] - F[3][i + N_1]|}{100(N_1 + N_2 - 2)} \right)$$

α refers to adjustable parameter and $\alpha = 0.45$.

If $N_1 + N_2 < N_f$, we continue to extract feature matrix of the next window and get its window length for comparison. Furthermore, *similarity* is assigned values according to the comparison results.

Step 2: Caculated *similarity* value is assigned to $\max S$ which is used to record the maximized *similarity* of all windows in the whole music file.

Step 3: Extract feature matrix of next window by order. Similar to the first algorithm, according to the comparison between extracted window length and the window length of piano fragments, similarity value can be respectively calculated.

Step 4: If $similarity > \max S$, *similarity* will be assigned to $\max S$. If there still has feature matrix in the window which cannot be extracted by order, return to step 3. Otherwise, if all the feature matrixes in the window of music file have been completed extracting, $\max S$ will be the similarity value of music file corresponding to piano fragments.

After all the music files have finished calculating its corresponding similarity value of piano fragments, some information of music file with the highest similarity value will be returned as retrieval results, based on the size of similarity values.

C. Similarity Matching of Humming Fragment Retrieval

During music retrieval based on humming fragments, similarity matching is an important step. First, melody feature information of music files is stored in feature database.

Second, the pitch period of each tone on humming fragments can be calculated by music retrieval algorithm of humming signal fragments.

On this basis, similarity matching computation needs to be performed and retrieval results should be returned. Since music retrieval algorithm of humming fragments can only compute pitch period of each tone on humming fragments, it cannot acquire determined melody feature matrix, the pitch period of each tone can only be applied. Meanwhile, by experiments and the means of fuzzy matching similarity algorithm of humming fragments, its similarity is calculated.

It is found in the experiments:

a) If there is difference in the pitch periods of two humming tones and it is less than four, that is, $|p_a - p_b| \leq 4$, $0 < a,b \leq N$. p is the pitch period of the tone and N is the number of tones in the humming fragment.

b) When $4 \leq |p_a - p_b| \leq 10$, $0 < a,b \leq N$, the two tones may be the same or neighbour ones, that is, $dis(P_a, P_b) \leq 1$. P is the value of tone.

c) When $|p_a - p_b| \geq 10$, $0 < a,b \leq N$, we can determine that two tones are not equal or adjacent.

For the calculation method of fuzzy matching similarity on humming fragments, each music file will perform fuzzy matching similarity calculation to get similarity value. Then, according to similarity sorting, the highest similarity music file information will be returned as retrieval results. Fuzzy matching similarity algorithm of each music file is calculated as follows

Step 1:Calculate two corresponding pitch changing arrays of the period of humming signal fragments. We set $C[i]$ as the array which stores the period of humming signal fragments, $C[i] = p_a, 0 < i \leq N_m$. N_m is the number of tones of humming signal fragment. Assume there is array $CF_2[i]$ ($0 < i \leq N_m$) to describe the changing relation between the pitch period of current tones and that of previous tone. So according to formula 5 we have

$$CF_2[i] = \begin{cases} 0 & i = 1 \\ C[i] - C[i-1] & 0 < i \leq N_m \end{cases} \quad (7)$$

Assume there is array $CF_3[i]$ to describe the changing relation between the pitch period of current tones and that of the first tone in humming signal fragments. According to formula 8 we have

$$CF_3[i] = \begin{cases} 0 & i = 1 \\ C[i] - C[l] & 0 < i \leq N_m \end{cases} \quad (8)$$

Step 2: Exact the first windowed feature matrix W_1 of the music file. If pitch number of humming signal fragments is the same to window length, following calculation can be carried out: Set $W_1[i][j]$ as window feature matrix and $0 < i \leq 3$, $0 < j \leq 20$. $N = W_1[l][l]$ refers to pitch numbers of humming fragments($N = N_m$, $2 \leq i \leq N$).

$$similarity = \frac{\sum_{i=2}^{N} CF_1(i)}{60(N-1)} + \frac{2\sum_{i=2}^{N} CF_2(i)}{100(N-1)} \quad (9)$$

Since the people tend to make correct comparison to the changing humming of neighbour two tones, the corresponding parameter of melody feature in similarity calculation will be higher and its value is 0. 6.

If window length is larger than pitch number of humming fragments, it means they are not matched

completely. So $similarity = 0$. If window length is smaller than pitch number of humming fragments, it indicates that this humming fragment is the music fragments with two or more than two sentences. This situation needs following judging calculation. Extract next window of feature matrix W_2. Set $N_1 = W_1[l][l]$ as the length of the first window and set $N_2 = W_2[l][l]$ as the length of the second window. N_m refers to pitch number of solved humming signal fragments. If $N_1 + N_2 > N_m$, it indicates they are not matched completely. Assign value of $similarity$ as 0. If $N_1 + N_2 = N_m$, we have:

$$CF_3[i] = \begin{cases} CF_3[i] & 1 < i \leq N_1 \\ CF_3[i] + C[l] - C[N_1 + 1] & N_1 < i \end{cases} \quad (10)$$

$$similarity = \frac{3(\sum_{i=2}^{N_1} SF_1(i) + \sum_{i=2}^{N_2} SF_1(i))}{50(N_1 + N_2 - 2)} + \frac{2(\sum_{i=2}^{N_1} SF_2(i) + \sum_{i=2}^{N_2} SF_1(i))}{50(N_1 + N_2 - 2)} \quad (11)$$

If $N_1 + N_2 < N_m$, we continue to exact feature matrix of next window and get window length for comparison. According to the results we assign value to $similarity$

Step 3: The value of calculated $similarity$ is assigned to $\max S$ which is used to record the maximized similarity of all windows in the whole music file.

Step 4: Exact feature matrix of next window based on order. Similar to the first step, according to the comparison between extracted window length and the window length of humming fragments, $similarity$ will be calculated respectively.

Step 5:If $similarity > \max S$, $similarity$ will be assigned to $\max S$. If there still exists feature matrix of window which has not been extracted based on sequence, return to step 4. If the windowed feature matrixes are all extracted, $\max S$ is the similarity value of this music file, corresponding to the humming signal fragments.

After all the music files finish calculating their corresponding humming fragments, we sort the music files according to the size of similarity value. Then, some information of music files with the highest similarity is returned as retrieval results.

IV.   EXPERIMENTAL ANALYSIS

There are 100 pop-song files and their eigenvectors in system database. The segments of randomly selected 20 songs are played by the piano to produce music segmentation files. Music files are retrieved through these 20 music segmentation files and there are 19 music segmentation files can be retrieved to corresponding music file with 1 wrong music retrieval. It means that three returned files don't exist in the file information of this music segmentation. There are totally 10 people taking part in this experiment without any training females and males. 88 humming fragments are recorded by means of singing words in quiet room and each

TABLE I.   VALUE OF $SF_1[i]$ AND $SF_1[i]$

$SF_1[i]$			$SF_2[i]$		
value	$W_1[2][i]$	Range of $CF_2[i]$	value	$W_1[2][i]$	Range of $CF_2[i]$
0	0	$(10, +\infty)$	0	0	$(10, +\infty)$
5	0	$[4,10]$	5	0	$[4,10]$
10	0	$(-\infty, 4)$	10	0	$(-\infty, 4)$
5	1	$(10, +\infty)$	5	1	$(10, +\infty)$
10	1	$[4,10]$	10	1	$[4,10]$
5	1	$(-\infty, 4)$	5	1	$(-\infty, 4)$
0	>1	$(10, +\infty)$	0	>1	$(10, +\infty)$
10	>1	$(-\infty, -10)$	10	>1	$(-\infty, -10)$
5	>1	$[4,10]$	5	>1	$[4,10]$
0	>1	$(-\infty, 4)$	0	>1	$(-\infty, 4)$
10	<-1	$(10, +\infty)$	10	<-1	$(10, +\infty)$
0	<-1	$(-\infty, -10)$	0	<-1	$(-\infty, -10)$
5	<-1	$[4,10]$	5	<-1	$[4,10]$
0	<-1	$(-\infty, 4)$	0	<-1	$(-\infty, 4)$

fragment is among 14-30s. Recording format is 44. 1KHz and 16bit quantification. There are totally 192 songs of MIDI format of Chinese popular music in music database. The system will sort retrieval results from large to small based on correlation and return the information of previous 10 songs with the highest correlation degree.

Audio processing needs to extract its attributes including non-performing attributes like file name, type, coded format and audio features with audio data. During extracting, data will be divided into frames. The features of each frame's tone and loudness will be extracted and combined. Audio is performed sampling training and classification to set up classified catalogues. Non-performing attribute information and feature information of extracted sound are totally stored in feature-oriented database. With records in meta-database and media association in audio database, users can retrieve audio information through their query interface. Users can inquire audio information or scan the classified catalogue. Towards long-fragment audio, users can begin scanning based on content, according to the audio structure. The retrieval engine retrieves required information of user with similarity instead of accurate match. The similarity between query vector and audio vector in the base is measured by distance. Each feature will have different distance measurement methods so as to be more effective in specific application.
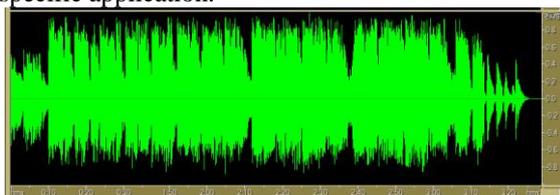

Figure 3.   Extracting feature of audio files.

Pitch of voice can be acquired by mainly local maximum amplitude of Fourier spectrum [20] and tone is the frequency of maximum amplitude. Frequency will be stored in a series of arrays with one dimension to be matched. We compare the recognition from different perspectives. The first mode is to randomly select half of labeled MIDI file as training set, while the other half of

this file is taken as test set. Note information through MIDI can be calculated to get eigenvectors of each melodic track. Eigenvectors in training set will be taken as input to respectively train these two kinds of classification models. After training, the accuracy of test classification model in test set will be applied and test results will be ecorded. The above process is repeated 10 times.

TABLE II.   COMPARISON OF EXTRACTED FEATURE FILES WITH THE FILES IN DATABASE

Feature files	Files in database
7002	5308
5968	7741
11774	10652
11865	13082
14005	18600
2386	2991
6940	6302
3861	8877
14118	14207
11205	10853
3088	3495
23599	19978
9743	6408
3842	4032
24289	23921
9956	10350
6655	5901
12991	10193
5234	6674
9958	7739
4848	7009
12110	14105

The accuracy rate of these two methods are shown by figure 4. From this figure, it is shown that average recognition accuracy rate of melodic track feature based on this paper's method is 84. 2%, while average recognition accuracy rate based on comparison method is 71. 3%.

If those songs with unclear melodies are deleted from MIDI files, a new dataset will be acquired. The so-called unclear melody refers that some songs don't have clear melody or the song melody can only be combined by many melodic tracks. Pop songs don't usually have this melodic format. On this basis, the experiment with the

same step of the first part is performed. These two methods' accuracy rate is shown as figure 5. From this figure, average recognition accurate rate of melodic track based on our method is 91. 7%, while average recognition accurate rate of comparison method is 73%.
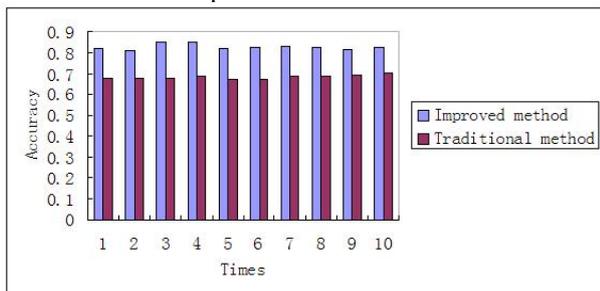

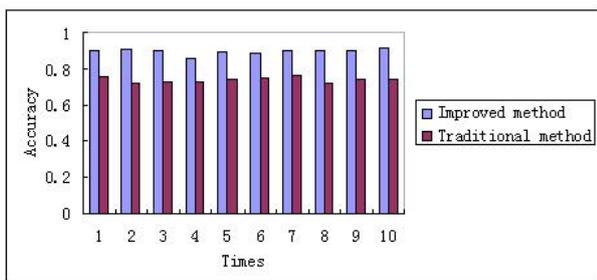Figure 4.   Comparison of recognition accuracy.

Figure 5.   Recognition rate in new data set.

Different from evaluation standards of comparative recall ratio and accuracy ratio on file retrieval, the evaluation standards of defining humming retrieval system is average query success rate and average retrieval time. Average query success rate refers to the probability to check users' humming songs appearing in returning results' list. Meanwhile, the sequence of this song in result list is checked. If its rank is in the front place, it indicates the performance is better. This paper has also improved the retrieval algorithm. Thus, under the precondition of correct retrieval results, average retrieval time will be taken as an evaluation standard. The test results are shown as figure 6 and 7.

From the experiments we can see that no matter users adopt humming pattern with lyrics or without lyrics, the system can have reasonable retrieval results. Retrieval success rates of top ten are all more than 80% and it proves the effectiveness of the algorithm in this paper. Meanwhile, it is seen that the effect of users' retrieval with lyrics humming is better than that without lyrics. It is mainly because users can perform change of tone height more correctly when they are humming with lyrics. However, if they are humming without lyrics, users are not very clear to the change of tone height so as to result in lowering retrieval success rate.

It is also found if the tone height can express melody, some songs with high similarity will appear in the retrieved results of list. It will cause that targeted songs cannot be correctly located. This phenomenon will be more and more serous with music database increase. That is to say, retrieval accuracy rate will decrease with enlarging music database. From figure 8 and 9, retrieval

accuracy rate of improved algorithm is improved to a large extent and it also proves the effectiveness of melody matching based on similarity.
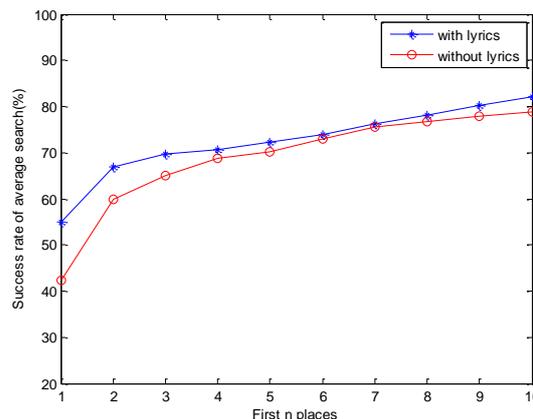
Figure 6.   Success rate of average search of two humming styles.
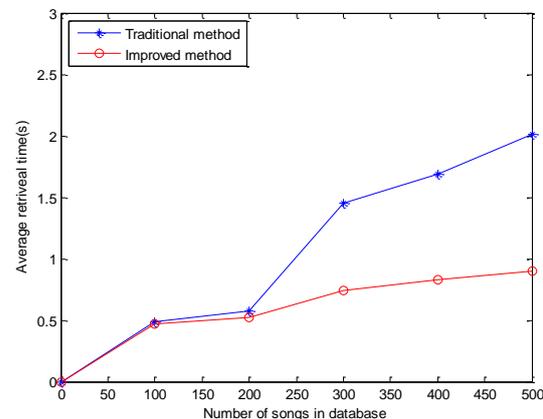
Figure 7.   Retrieval speed comparison of two methods.
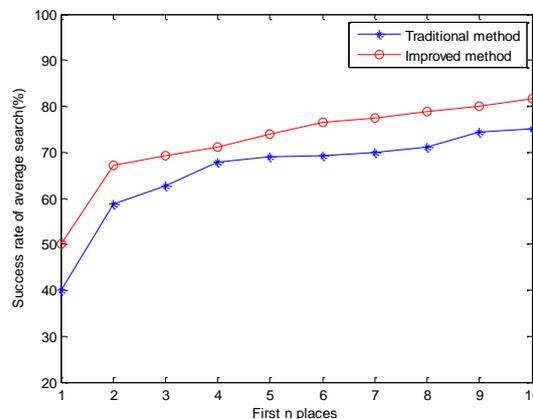
Figure 8.   Success rate of average search of two methods with lyrics

V.   CONCLUSION

With constant development of network and multimedia technology, it becomes more and richer to describe information with video and audio in network. The retrieval needs of networked information are also become stronger. However, most current retrieval engine is to

retrieve in the whole version and the research of multi-media retrieval is In its infancy . Music retrieval theories based on content at home and overseas develop very fast. However, due to high complex algorithm, slow retrieval speed and narrow application, the domestic researches are lacked. Thus, this paper studies the key technology of music retrieval on the basis of similarity match and mainly completes the following two jobs. One of them is the retrieval algorithm research based on music content. The latest music retrieval algorithm based on content is performed detailed description and analysis. Unsteady length segmentation algorithm based on music fragments is mainly analyzed by us. Furthermore, the calculation method of accurate matching similarity corresponding to music fragment retrieval and that of fuzzy match similarity corresponding to humming fragment retrieval are proposed. From the experiments, compared to traditional music fragment retrieval, the method in this paper effectively improves matching result accuracy and it achieves good performance on query processing speed. This paper's task is to lay a good foundation of music retrieval based on content and it has promoted and referential significance for further research.
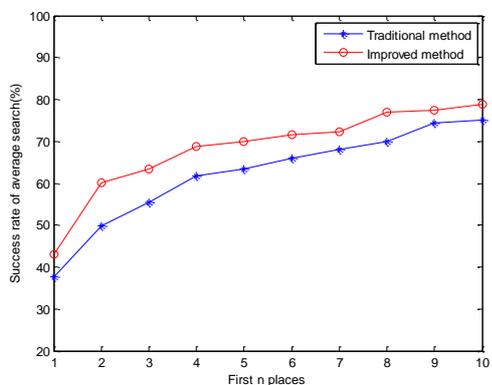

Figure 9. Success rate of average search of two methods without lyrics

REFERENCES

[1] A J Ghias, Logan D Chamberlain, B C Smith, "Query by humming-musical information retrieval in an audio database", *Proceeding of the ACM Multimedia' 95*, pp. 231-236, 1995.

[2] R J McNab, L A Smith, I H Witten, "Towards the digital music library: Tune retrieval from acoustic input", *Proceeding of the ACM Digital Libraries Conference*, pp. 11-18, 1996.

[3] Jin Yi, Huang Min, "Study on the Melody Based Retrieval of Music—Representation and Extraction of the Melodic Features", *New Technology of Library and Information Service*, vol. 18, no. 4, pp. 23-26, 2003.

[4] Zhenjing Yao, Tao Gao, Guo Li, "The Optimized Pseudorandom Digital Modulation Excitation Sequences for Multichannel Ultrasonic Ranging System", *Journal of Networks*, vol. 7, no. 10, pp. 1600-1609, 2012.

[5] Nicola Montecchio, Emanuele Di Buccio, Nicola Orio, "An Efficient Identification Methodology for Improved Access to Music Heritage Collections", *Journal of Multimedia*, vol. 7, no. 2, pp. 145-159, 2012.

[6] R. J. McNab, L. A. Smith, I. H. Witten, et al, "Towards the digital music library: tune retrieval from acoustic input", *Proceedings of Digital Libraries*, pp. 11-18, 1996.

[7] Rho Seungmin, Hwang Eenjun, "FMF:Query adaptive melody retrieval system", *Journal of Systems and Software*, vol. 79, no. 1, pp. 43-56, 2006.

[8] Marolt Matija, "A mid-level representation for melody-based retrieval in audio collections, IEEE Transactions on Multimedia, vol. 10, no. 8, pp. 1617-1625, 2008.

[9] Chang Chuan-Wang, Jiau Hewijin Christine, "Numeric indexing and access mechanism for melody retrieval, International Journal of Innovative Computing", *Information and Control,* vol. 7, no. 7, pp. 4083-4096, 2011.

[10] C. L. Parker, "A tree-based methed for fast meledic retrieval", *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, pp. 254-255, 2004.

[11] LIU Yi, HAO Yun-fei, "Evaluation of approximate match algorithms in large query by humming music retrieval systems", *Journal of Hunan University of Science & Technology*, vol. 17, no. 11, pp. 810-816, 2009.

[12] Shen Zhou, Wang Yongcheng, Liu Gongshen, "Efficient Skip-Pattern Matching Algorithm for Approximate String Sequential Problem", *Journal of Data Acquisition & Processing*, vol. 8, no, 2, pp. 56-60, 2001.

[13] Yu-Lung Lo, Wen-Ling Lee, Lin-huang Chang, "True suffix tree approach for discovering non-trivial repeating patterns in a music object, Multimedia Tools and Applications, vol. 39, no. 2, pp. 169-187, 2007.

[14] Huawang Shi, "Integration of Unascertained Method with Neural Networks and Its Application", *Journal of Networks*, vol. 6, no. 11, pp. 1631-1639, 2011.

[15] JI Xing-an, "Study of query by humming based on ant colony DTW algorithm", *Information Technology*, vol. 23, no. 9, pp. 236-239, 2011.

[16] SUN Jinli, CHEN Jie, DENG Min, "The Algorithms of Geometry Similarity Measurement and Experimental Analysis for Linear Spatial Data Transmission", *Journal of Geo-Information Science*, vol. 17, no. 5, pp. 334-342, 2011.

[17] Suzuki Motoyuki, Hosoya Toru, Ito Akinori, "Music information retrieval from a singing voice using lyrics and melody information", *Proceeding of Eurasip Journal on Advances in Signal Processing*, pp. 200-207, 2007.

[18] Ellis Daniel P. W. , Poliner Graham E, "Classification-based melody transcription", *Machine Learning*, vol. 65, no. 2, pp. 439-456, 2006.

[19] Wu Ya-Dong, Li Yang, Liu Bao-Long, A new method for approximate melody matching", *International Conference on Machine Learning and Cybernetics*, vol. 5, pp. 2687-2691, 2003.

[20] HU Tai-ping, JI Yun-gang, PAN Yang-yong, "Application of Fourier Spectrum Decomposition Technique to Hydrocarbon Exploration in TZ-45 Wellblock in Tarim Basin", *Xinjiang Petroleum Geology*, vol. 11, no. 3, pp. 45-49, 2011.

Research on Segmentation of Pear Shape from Unorganized Point Clouds

Hui-jun YANG[1, 2]

[1]College of Mechanical and Electronic Engineering, Northwest A&F University, Yang ling, Shaan xi, China
[2]College of Information Engineering, Northwest A&F University, Yang ling, Shaan xi, China
Email: yhj740225@nwsuaf.edu.cn

Dong-jian HE

College of Mechanical and Electronic Engineering, Northwest A&F University, Yang ling, Shaan xi, China
Email: hdj168@nwsuaf.edu.cn

Linhao Li

AT&T Labs, 200 South Laurel Ave, #D4-3C05, Middletown, NJ, USA
Email: lli@research.att.com

Shao-Hua JIANG

Institute of Image Recognition & Computer Vision, Hunan Normal University, Changsha, Hunan, China
Email: jiangshaohua@smail.hust.edu.cn

Abstract—With the advent of 3D scanner, accurate segmentation of 3D fruit shape from unorganized point clouds has been turned out to be the most challenging task in scientists and engineers in reverse engineering. This paper herein proposes efficient and robust approach to extract pear shape from background. At first, an interactive, non-local denoising algorithm is employed to efficient denoise the pear scans; Second, geometric properties, including normal, variation and curvature, are estimated by covariance analysis; Third, a Recursive Region Increment (RRI) is proposed to add the geometric similarity points to a base set, to generate an ultimate set only including the points of pear shape; Forth, point clouds is linearized for rapidly rending in the post processing. Finally, segmentation algorithm applied on ten range scans of a pear demonstrates that our algorithm reduces the number of pear point clouds by 88.3%, proves the validity and practicability of this method in pear segmentation.

Index Terms—Denoising, Covariance Analysis, Segmentation, Linearization, Point Cloud

I. INTRODUCTION

A. Related Work

In recent years, with the development of 3D scanning technology [1] [2], rich details of the object shape can be acquired with dense sampling points (point cloud). Thus, 3D scanning of real objects has been popularly used for information acquisition in various applications. From then on, most research has focused on the reconstruction of 3D object from point clouds in 3D city modeling [3], human face reconstruction [4], architecture reconstruction [5], and tree reconstruction [6], so it's probable to use point clouds in fruit reconstruction.

On the other hand, the mesh representation and image processing is still used for digitalization of real fruits [7], even though there is a growing demand for more detailed and real models in agriculture. So in recent years, G. W. Kim [8] has begun to set about the actual apple model through finite element method (FEM) simulation, but it takes a long time to render highly detailed meshes with more artifacts. To overcome this drawback, new processing methods must be presented to preprocess, reconstruct and render these highly complex geometric bodies. Thus in 2009, 3D scanner has been first adopted to acquire 3D digital primitives—rending primitives of irregular shaped food by Uyar R [9]. In contrast to the mesh representation, due to not requiring connectivity information for describing topological relations between vertices, point images can provide relatively fast visualization of huge 3D models and make it more realistic to obtain accurate results from simulation. So it's necessary to reconstruct pear shape from point clouds.

From above analysis, we discover that construction of pear shape from point cloud is more important, but the research on it is not intense. The main reasons are that, in those applications, the points can be obtained without any background; while the pears are often grown on trees with leaves, trunks and branches, so the pear point clouds are unorganized, noised and mixed, which makes it hardly used for pear shape reconstruction in agriculture. As a result, denoising and accurate pear shape segmentation becomes an important topic in pear shape

---

corresponding author: Dong-jian HE, affiliations: College of Mechanical and Electronic Engineering, Northwest A&F University, No.3 Taicheng Road, Yang ling, Shaan xi, China. Postcode: 712100

reconstruction, and they must be done before classical process.

The current classic denoising algorithms are divided into the following four main categories: partial differential equation (PDE) [10], spectrum technology [11], statistical techniques [12] and bilateral filtering algorithm [13]. These algorithms are possible to achieve highly conformal effect but have low efficiency, large computation and can't fit for different parts of pear shape. So, it's necessary to present an efficient denoising algorithm for the different parts of pear. The current segmentation methods mainly focused on building features [3], characteristic edges [17], valley-ridge [18] or curve skeleton [19]. All these state-of-the-arts can only segment a specific geometric shape, but can't be used for segmentation of the random pear shape in our task, so the previous methods will not be feasible.

B. Contributions and Outline

In this paper, an interactive testing algorithm is extended for efficient denoising of different parts of pear; Covariance analysis is employed for estimation of pear geometric feature; Recursive Region Increment (RRI) process is proposed for extracting random pear points from massive point cloud; Linearization is used for efficient storage, rendering and out-of-core representation.

This paper is organized as follows. Section 2 describes the proposed algorithm; Section 3 presents simulation results of the presented algorithm on the pear shape; Section 4 concludes this paper and discusses future works which will make this work more perfect.

II. PROPOSED ALGORITHM

Point clouds from laser scanners are a set of spatial points in a three dimensional Cartesian coordinate system as in (1)

$$R = \{x_i, y_i, z_i\} \ (i = 1, 2 \ldots K) \qquad (1)$$

It's really unorganized 3D points set and there's no topological relationship among them. The extraction is a process of segmenting the geometric-similarity points from original point clouds and the extracted subset is geometrical significance points set, each set is a piece of complete pear scan.

In this work, a four-step algorithm is presented to segment pear shape from unorganized point clouds. At first, the denoising process is employed to preprocess unorganized point clouds; Second, the classical methods covariance analysis are employed to estimate the normals, variance, curvature and tangent vector of each point; Third, starting from a seed point, an segmentation algorithm is presented to extract ultimate purely pear points set from leaves and background by searching surrounding points recursively until encountering the border of leaves and fruit, in each iterative step, the estimated features are used to evaluate geometric similarity of two points; Fourth, extracted points are organized into a linear memory structure for more

quickly access. The pipeline of this step is shown in Fig. 1.

A. Denoising Process

Denoising is a key task in point clouds processing, and many classic algorithms has been presented by now. Early method is summarized as the following four categories: The first category is partial differential equation (PDE) [10] that can extend curvature flow to the denoising process later. The second category is spectrum technology proposed by Pauly et al. [11], which create a spectral decomposition for point cloud, and deal with the noise by handling spectrum coefficient. The third category is based on statistical data analysis framework presented by Pauly et al. [12]. The last category is to extend the famous image bilateral filtering algorithm to mesh model for non-local neighborhood denoising [13] by calculating a local radial basis function (RBF) approximation. These algorithms have low efficiency and large amount of calculation, thought it is possible to achieve highly conformal effect.
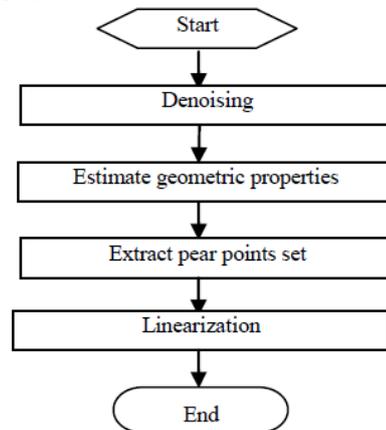
Figure 1.   Pipeline of extraction algorithm for pear.

In this paper, based on interactive testing proposed by Weyrich et al [14], outliers on pear point clouds are divided into 2 types by heuristic, user-specified method. Type 1 is the point clouds of the surfaces around pear head and stem, which appear rich local details and smooth surface. For these surfaces, global similarity is first measured by weighted average of similar points within neighborhood [15], and then the smallest enclosing ball approach is adopted to denoise local rich details [14]. Type 2 is the point clouds of the pear side surfaces, which appear smooth surface and boundary characteristics. For these surfaces, outliers are classified by taking weighted average of similar points within neighborhood as measurement of nearest neighbor interaction [15], then mean-shift iterations is employed for denoising by drifting each sample point to the local maximum of density function based on selection of different surface fitting and kernel function. By applying non-local denoising method, the process of local approximation is omitted, estimation speed is more rapid, and finally the denoising process is implemented speedily without filtering, as shown in Fig. 2.

Based on the statistical weights and nonlocal similar measurements of mean-shift, the follow-up estimation of all geometry properties will be more precise.

B. Estimation of Geometric Properties

Geometric properties estimation is a key step and important foundation for extracting pear shape. By now, each point has only two properties: position and initial normals. On the other hand, edges and surfaces of pear scans can be considered as curves and curved surface respectively, while leaves consist of a lot of singular and uneven points, which results in great variance on properties of leaves' points. So, more geometric features, including precise normals, variance and curvature, should be estimated by the covariance analysis due to its merit in finding geometric features without losing much information [16].
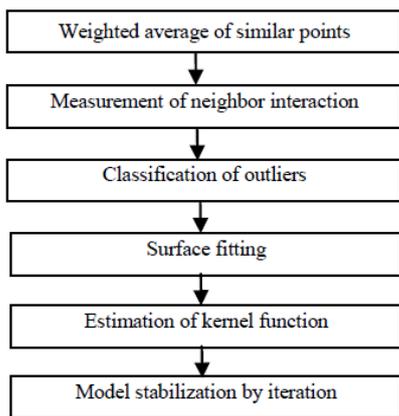
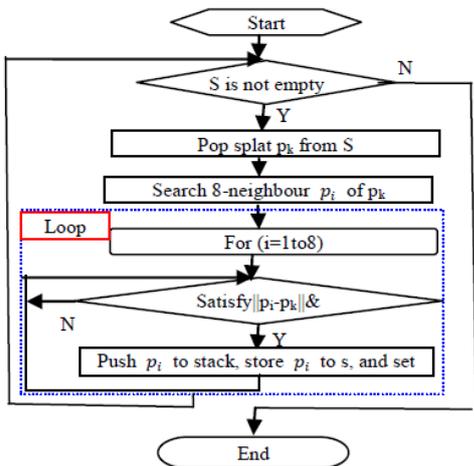Figure 2.   Denoising of pear side face point cloud.

Figure 3.   Pipeline of extracting pear point set from all point cloud.

A unit normals, in this paper, can be considered as point on a unit sphere, and point normals n_i (i=0,...,m−1) can be expressed by its angles $n_i = (\theta_i, \varphi_i)$, $\theta \in [0, \pi]$, $\varphi \in (-\pi, \pi]$).

First, the mean normal n is computed by weighted average on spheres based on least squares minimization that reflects spherical distances. The second step is calculation of the covariance matrix. Covariance is usually measured in two dimensions, whereas we have a set of 3-dimensional points p_i, so, 9 covariance values need to be calculated and put in a 3*3 covariance matrix M in (2) [16].

$$M = \sum_{i=1}^{m} (p_i - p) \times (p_i - p) \qquad (2)$$

where, p is the center of points p_i.

The third step is calculation of the eigenvalue λ_i and corresponding eigenvector $X_i (i = 1,2,3)$ of matrix M in (3).

$$M \cdot X_i = \lambda_i \cdot X_i \quad \lambda_1 \leq \lambda_2 \leq \lambda_3 \qquad (3)$$

In keeping the standardization, all eigenvectors are scaled to unit vectors in (4).

$$X_i = X_i / \|X_i\| \qquad (4)$$

The eigen analysis gives us the variances of the Gaussian distribution. Estimated normals n at p is the unitized eigenvector of X_1, associated with the smallest eigenvalue λ_1, and surface variance V_i of the points set is given in (5)

$$V_i = \lambda_1 / (\lambda_1 + \lambda_2 + \lambda_3) \qquad (5)$$

where λ_1 indicates the sum of square distance from all points p_i to the local plane, $\lambda_1 + \lambda_2 + \lambda_3$ is the sum of square distance from all points p_i to the barycenter p. So the surface variance reflects the deviation degree of the points to the tangent plane of the points. In the flat region, the points with small V_i are close to same plane, while in the rough region, the points with large V_i are disperse, thus V_i is a good approximation of the curvature at p. So far, each point has three estimated properties, they are position p, normals n_i and variance (curvature) V_i, which will be used to extract pear in the following step.

C. Segmentation of Pear Point Set

Recently, research on the extraction of characteristic has become more and more popular, region growing algorithm was usually applied to extracting building features [3]; Philippe Marin et al [17] proposed a method for the extraction of characteristic edges detected with a prior parabolic fitting; Xu-fang Pang [18] et al proposed a robust algorithm for valley-ridge extraction from point set; Jun-jie Cao [19] et al presented an algorithm for curve skeleton extraction via Laplacian-based contraction particularly from point clouds; All these extraction methods have extracted a specific geometric shape. While in our task, the fruit shape is random, so the previous methods will not be feasible.

Our segmentation method is a Recursive Region Increment (RRI) process, which takes variation V_i and dot product of normals n_i as the segmentation parameter, stack S as temporary memory and the spatial point set C as the ultimate points. The extraction algorithm starts from a seed point p_k, which is selected from pear scans $\{p_i\}$ initialized by $flag_i = 0$. Then p_k is added to C to form a new set and pushed to the stack S, set $flag_k = 1$. When the stack is not empty, repeat the follow three-step:

(1) Pop a point p_k from stack S;

(2) Seek the 8-connected neighbor point p_i (i=1, 2…8) around p_k in xyz directions by calculating the Euclidean
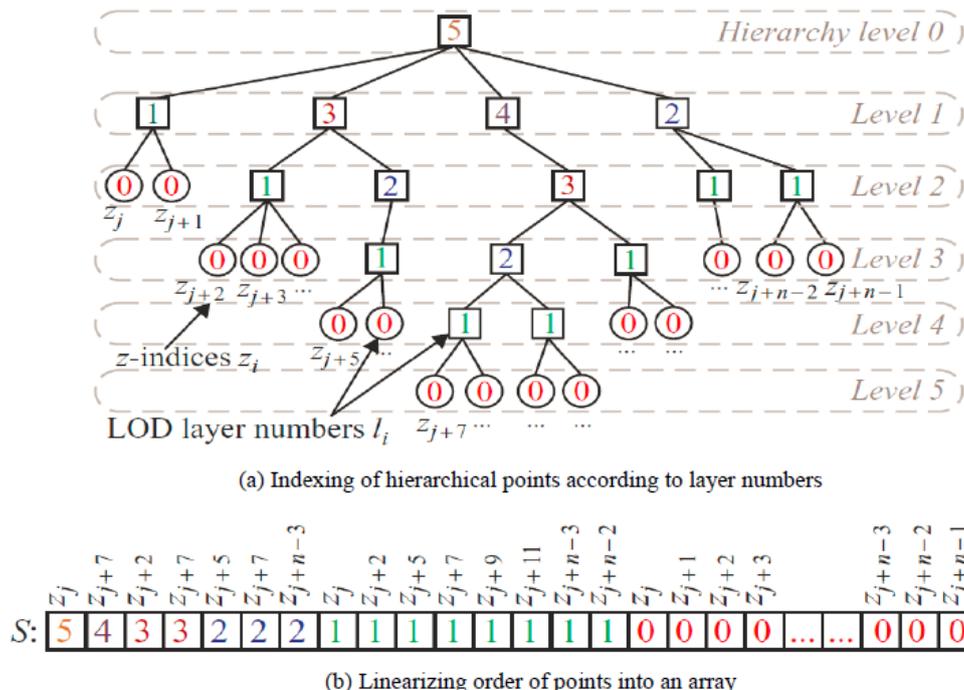
(a) Indexing of hierarchical points according to layer numbers

(b) Linearizing order of points into an array

Figure 4.   Linearization of hierarchical structure.

TABLE I.   PARAMETERS OF PEAR SCANS

items / scans	X extent (min,max)/mm	Y extent (min,max)/mm	Z extent (min,max)/mm	size (x*y*z)/mm	Num of points
Bottom1	(-32.46,62.78)	(-110.23,74.11)	(-84.2,75.13)	95.24*184.34*159.33	175583
Bottom2	(-31.93,63.23)	(-109.26,73.4)	(-85.02,73.16)	95.16*182.66*158.18	171982
Top1	(-33.34,62.14)	(-111.4,71.51)	(-83.67,74.19)	95.48*182.91*157.86	184350
Top2	(-34.27,62.18)	(-110.2,72.63)	(-82.31,75.15)	96.45*182.91*157.46	184319
Side1	(-33.81,63.56)	(-109.6,73.14)	(-83.58,73.41)	97.37*182.74*156.99	190024
Side2	(-35.5,63.62)	(-111.21,72.34)	(-84.14,74.57)	99.12*183.55*158.71	196543
Side3	(-35.28,62.74)	(-110.2,74.14)	(-85.32,75.7)	98.02*184.34*161.2	188572
Side4	(-32.34,63.25)	(-109.98,71.88)	(-82.35, 74.65)	94.59*181.86*157	190631
Side5	(-34.23,62.46)	(-111.2,73.26)	(-83.1, 73.89)	96.69*184.46*156.99	191778
Side6	(-33.54,63.28)	(-109.17,72.21)	(-82.4,74.13)	96.82*181.38*156.53	191253

distance between p_k and the 8-connected points, with $flag_i = 0$.

(3) For each p_i, the dot product of normals $|n_i \cdot n_k|$, $(\|n_j\| = 1)$ and surface variance difference $|V_i - V_k|$ are used as geometric similarity $\|p_i - p_k\|$ of all leaves and the pear. When the dot product satisfies $|n_i \cdot n_k| > 0.9$, points p_i are on the same surface with p_k. $|V_i - V_k|$ is another similar criteria of surface: lower difference value means p_i on the surface of pear; while higher difference value means p_i on leaves, so when $|V_i - V_k| > 0.025$, the point p_i can be treated as the edge of fruit and leaves. When geometric similarity $\|p_i - p_k\|$ are satisfied and $flag_i = 0$, push p_i to stack S, set $flag_i = 1$ and store p_i to C, else, return to step (3) for next p_i.

This three-step iteration will stop until stack S is empty, as shown in Fig. 3.

D. Linearization of Hierarchical Structure

Any spatial point representation, such as octrees or k-d trees [21] can be converted into a linear memory array and stored on disk.

In this paper, spatial structure will first be organized into linear structure for efficient storage and rendering, then an out-of-core representation would be obtained by mapping this linear memory file to disk, which allows our program to access array points directly by virtual memory (VM) manager that automatically decides which parts of the file should be loaded into main memory by demanding paging. Thus, all LOD points are lexicographically reordered respective to a layer number and a spatial ordering index, as illustrated in Fig. 4 [20].

III.   SIMULATION RESULTS

The point clouds of pear and leaves are scanned from 10 different angles by 3D Camega PCP-400, 2 range scans are bottom, 2 scans are top and 6 scans are side faces. These faces were managed by an array owned 10 elements in (6)

$$pear(10) = \{b_1, b_2, t_1, t_2, s_1, s_2, s_3, s_4, s_5, s_6\} \quad (6)$$

The original pear scans are shown in Fig. 5, those parts circled by red color is "pure" pear faces.
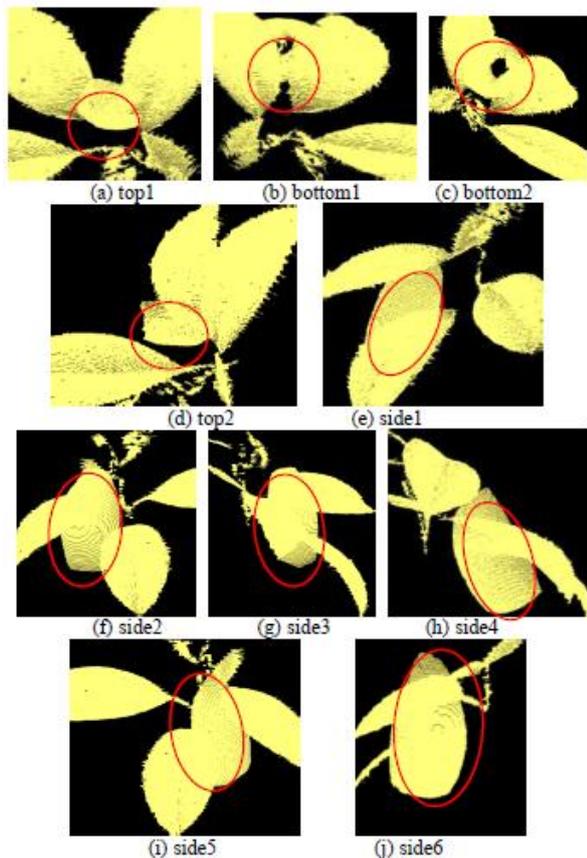
(a) top1   (b) bottom1   (c) bottom2

(d) top2   (e) side1

(f) side2   (g) side3   (h) side4

(i) side5   (j) side6

Figure 5.   The original point clouds of pear.

(a) top1   (b) bottom1   (c) bottom2

(d) top2   (e) side1

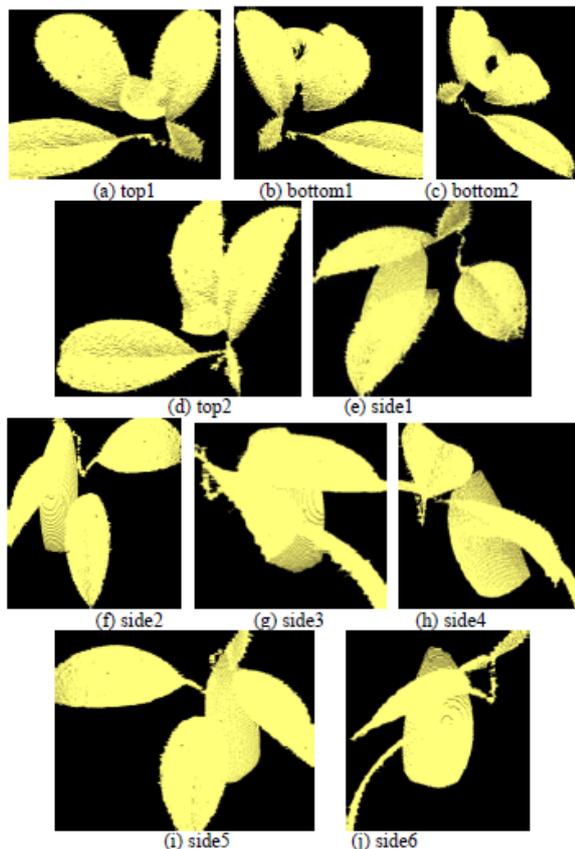(f) side2   (g) side3   (h) side4

(i) side5   (j) side6

Figure 6.   The denoised point clouds of pear.

The spatial resolution of point clouds is 0.03mm, and parameters of scans are shown in table1.

```
solid  pear(1)
 facet normal  -0.509082 -0.858324 0.064153
    outer loop
  vertex -28.810467 -5.502658 15.463493
  vertex -28.850819 -5.496768 15.222094
  vertex -28.517298 -5.709276 15.025510
    endloop
 endfacet
 facet normal  -0.499004 -0.862360 0.085615
    outer loop
  vertex -28.517298 -5.709276 15.025510
  vertex -28.850819 -5.496768 15.222094
  vertex -28.566435 -5.704361 14.788620
    endloop
 endfacet
 facet normal  -0.517902 -0.852686 0.068583
    outer loop
  vertex -28.850819 -5.496768 15.222094
  vertex -28.892342 -5.490918 14.981265
  vertex -28.566435 -5.704361 14.788620
    endloop
 endfacet
 facet normal  -0.515060 -0.853891 0.074726
    outer loop
  vertex -28.566435 -5.704361 14.788620
  vertex -28.892342 -5.490918 14.981265
  vertex -28.936321 -5.485356 14.741685
    endloop
 endfacet
 facet normal  -0.454053 -0.871978 0.183002
```

(a) Geometric features of bottom1

```
solid  pear(5)
 facet normal  -0.461369 -0.871547 -0.165967
    outer loop
  vertex -25.278475 -5.517074 -4.458734
  vertex -25.240240 -5.494243 -4.684917
  vertex -24.832659 -5.720109 -4.631847
    endloop
 endfacet
 facet normal  -0.446546 -0.862092 -0.239569
    outer loop
  vertex -24.832659 -5.720109 -4.631847
  vertex -25.240240 -5.494243 -4.684917
  vertex -24.755873 -5.698848 -4.851480
    endloop
 endfacet
 facet normal  -0.440534 -0.873202 -0.208441
    outer loop
  vertex -25.240240 -5.494243 -4.684917
  vertex -25.179163 -5.472042 -4.907007
  vertex -24.755873 -5.698848 -4.851480
    endloop
 endfacet
 facet normal  -0.432100 -0.867105 -0.247828
    outer loop
  vertex -24.755873 -5.698848 -4.851480
  vertex -25.179163 -5.472042 -4.907007
  vertex -25.098524 -5.449770 -5.125527
    endloop
 endfacet
 facet normal  -0.476030 -0.833404 -0.280774
```

(b) Geometric features of side1

Figure 7.   Estimated geometric features of pear

A. Denoising

At first, pear point cloud is denoised by the flowchart in Fig. 2 based on the programming in VC, point clouds in Fig. 5 (a) - Fig. 5(b) are dealt with methods proposed by Ref. [15], point clouds in Fig. 5(e)- Fig. 5(j) are dealt with mean-shift iterative methods proposed by Ref. [15], the denoised result is shown in Fig. 6.

After denoising, the point number of pear(10) is changed into the following list: {168427, 164826, 177194, 177163, 182868, 189387, 181416, 183475, 184622, 184097}.

B. Estimation Result of Geometric Properties

Then, every point is normalized by $\|n_i\| = 1$, normal and curvature of every point is estimated by covariance

TABLE II.   CONTRAST ON POINTS NUMBER OF PEAR SCANS BEFORE AND AFTER SEGMENTATION

scans	Bottom1	Bottom2	Top1	Top2	Side1	Side2	Side3	Side4	Side5	Side6
un-extracted	175583	171982	184350	184319	190024	196543	188572	190631	191778	191253
extracted	11186	7585	19953	19922	25627	32146	24175	26234	27381	26856
decrement rate /%	93.6	95.6	89.2	89.2	86.5	83.6	87.2	86.2	85.7	86

analysis. These estimated features of bottom1 and side1 of pear are shown in Fig. 7 respectively.

Fig. 7 gives part of the estimation results of geometric properties only in the form of text file. Among them, the content between facet and end-facet is the attributes of a point; The three values behind the facet-normal are the estimated normal vector of the current point; The content between outer-loop and end-loop is the center of neighbor points and surface information, in which, the three values at the back of vertex are the center of neighbor points and the three values behind f are the point surface change rate along three dimensions.

C. Extraction Result of Pear

Next, based on the extraction parameters in sec *II.C*, the pear surfaces are final segmented from leaves, the ultimate segmented pear range scans are shown in Fig. 8.

(a) top1   (b) top2

(c) bottom1   (d) bottom2   (e) side1

(f) side2   (g) side3   (h) side4
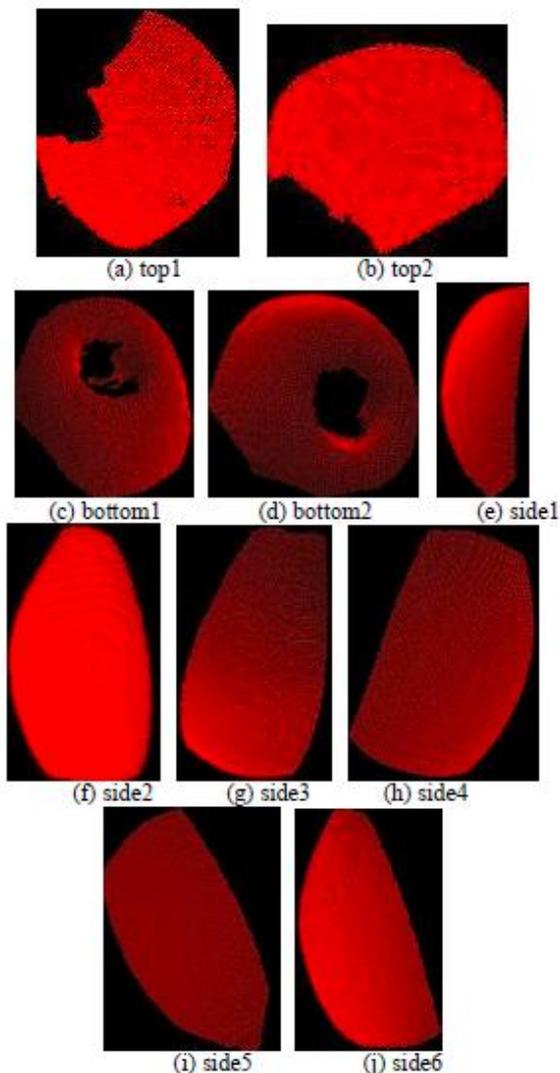
(i) side5   (j) side6

Figure 8.   The extracted pear scans

D. Comparison

After extraction, the number of points is markedly reduced; the contrast on number between un-segmented and segmented of all pear scans is presented in table 2.

Analysis of data from table 2 shows that, point number of segmented pear scans is decreased by 88.3% on average than those of un- segmented. This is a down of larger proportion, due to the larger proportion of leaf and branch in point clouds, so the number of points decreases obviously after segmentation.

IV.   CONCLUSION AND FUTURE WORK

In this paper, an efficient strategy is proposed to segmenting pear from unorganized background. An interactive, non-local and mean-shift based denoising algorithm is employed for efficiently removing the outliers; covariance analysis is presented for precise estimation of geometric features; a Recursive Region Increment (RRI) process is proposed to segment pear shape from background speedily, in sharp contrast to the traditional cluster methods, those points outside the bounding needn't to be processed, which minimizes the number of processed points and makes the segmentation process rapidly. Additionally, spatial structure of points was linearized for the efficient access. The algorithm has reduced the calculation complexity, speed up the extraction process while remaining details of pear shape, reduce the number of points by 88.3%, and can be extended to segmentation of other fruits shape with smooth surface. Even though, this work is only suitable for segmenting continuous smooth surface, and can't be used for coarse surface.

REFERENCES

[1] N. A. Borghese, G. Ferrigno, G. Baroni, et al. Autoscan: A Flexible and Portable 3D Scanner, IEEE Computer Graphics and Applications , 1998:18 (3), pp. 38-41.

[2] B. Koyuncu, Kurtulus Kullu. Development of an optical 3D scanner based on structured light, Proceedings of the 9th WSEAS international conference, 2010, pp. 17-22.

[3] G. Sithole, G. Vosselman, "Automatic structure detection in a point cloud of an urban landscape", Proceedings 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and

Data Fusion over Urban Areas, URBAN2003, Berlin, Germany, pp. 67-71

[4] C. J. Zhou, X. P. Wei, Q. Zhang and B. X. Xiao, Image Reconstruction for Face Recognition Based on Fast ICA, International Journal of Innovative Computing, Information and Control, vol.4, no.7, pp. 1723-1732, 2008

[5] L. L. Nan, A. Sharf, H. Zhang, et al. Smart Boxes for interactive urban reconstruction. Siggraph2010 Proceedings, 2010: 29(4), pp. 93:1-93:10.

[6] Rosell, Joan R, L. Jordi, S. Ricardo, et al. Obtaining the three-dimensional structure of tree orchards from remote 2D terrestrial LIDAR scanning , Agricultural and Forest Meteorology, 2009: 149(9), pp. 1505-1515.

[7] L. Lleo, P. Barreiro, M. Ruiz-Altisent, et al. Investigation of the effect of shape on the acoustic response of 'conference' pears by finite element modeling , Journal of Food Engineering, 2009: 93(2), pp. 229-235.

[8] G. W. Kim, G. S. Do, Y. Bae, et al. Analysis of mechanical properties of whole apple using finite element method based on three-dimensional real geometry, Food Science and Technology Research, 2008:14(4), pp. 329-336.

[9] R. Uyar, F. Erdogd. Potential use of 3-dimensional scanners for food process modeling, Journal of Food Engineering, 2009:93(3), pp. 337-343.

[10] P. Perona, J. Malik. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1990: 12(7) pp. 629-639.

[11] M. Pauly, M. Gross. Spectral processing of point-sampled geometry. Computer Graphics, 2001:35(4) pp. 379-386.

[12] M. Pauly, N. Mitra, L. Guibas. Uncertainty and Variability in Point Cloud Surface Data, Eurographics Symposium on Point-Based Graphics2004, 2004 pp. 77-84.

[13] S. Yoshizawa, A. Belyaev, H. P. Seidel. Smoothing by example: Mesh denoising by averaging with similarity-based weights. In Proceedings of Shape Modeling International, 2006 pp. 38-44.

[14] T. Weyrich, Mark Pauly, R. Keiser, S. Heinzle, S. Scandella, and M. Gross. Post-processing of scanned 3D surface data. In Proceedings of Eurographics Symposium on Point-based Graphics 2004, 2004 pp. 85-94.

[15] O. Schall, A. Belyaev, H. Seidel. Feature-preserving Non-local Denoising of Static and Time-varying Range Data. Computer-Aided Design, 2008: 40(6) pp. 701-707.

[16] Smith LI. A tutorial on principal components analysis. Maintained by Cornell University, USA; 2002.

[17] P. Marin, A. Meyer, V. Guigue. Partition along characteristic edges of a digitized point cloud. Proceedings of the international conference on shape modeling & applications. 2001 pp. 326.

[18] Xu-Fang Pang, Ming-Yong Pang. An Algorithm for Extracting Geometric Features from Point Cloud. Proceedings of the 2009 International Conference on Information Management, Innovation Management and Industrial Engineering. 2009, pp. 78-83.

[19] Junjie Cao, Andrea Tagliasacchi, Matt Olson, et al. Point Cloud Skeletons via Laplacian Based Contraction. Proceedings of the 2010 Shape Modeling International Conference. 2010, pp. 187-197.

[20] M. Gross, H. Pfister, et al. Point-Based Graphics. Morgan Kaufmann Publishers. 2007. pp. 149-166, 109-127, 408-412.

[21] K. Zhou, Q. M. Hou, R.Wang.et al. Real-time KD-tree construction on graphics hardware. ACM Transactions on Graphics, 2008:27(5), pp. 126:1-126:11.

Huijun YANG was born in Wanrong, Shanxi, China in Feb 1974. She received her M.Sc. degree in Computer Science Technology at National University of Defense Technology, Changsha, China in 2005 and the B.S. degree in Shanxi University, Shanxi, China in 1997. Her current major field of study includes computer graphics, point cloud reconstruction and E-Commerce.

She has been the lecture of the College of Information Engineering, Northwest A&F University, in 2002-2010, has been a senior visiting research fellow in Nanjing University, Nanjing, China, in 2011. She is now working as vice professor in the College of Information Engineering, Northwest A&F University in Yangling, Shaanxi, China. Hitherto, she has published over 18 academic papers in the significant journal, about 6 articles of them has been cited by EI index. She has presided 3 projects and participated 6 projects as the key researcher. She has also published 2 textbooks as the paternal. Three of the main achievements are:

A simplified and accurate registration for splat-based fruit scans. ICIC Express Letters, Part B: Application. 2012, 3(4):733-741.

Optimal pairwise registration of fruit scans based on splat-oriented. International Journal of Advancements in Computing Technology. 2012, 3(4): 148-159.

A novel algorithm for segmenting fruit from unorganized point clouds. VRCAI2011: ACM SIGGRAPH Conf.
Currently, she is interested in fruit shape reconstruction in agriculture, point cloud segmentation and graphics technology.

Dr. Huijun YANG won Grand Prize of paper presentation on 2012 National Agricultural Engineering doctoral Forum in 2012, 1 second prize of Excellent Paper on doctoral Forum in northwest A& F University in 2012, 1 Excellent Paper Award in northwest A& F University in 2007, 1 third prize in Young Teachers Lecture Competitions in northwest A& F University in 2009.

Dongjian HE was born in Hanzhong, Shaanxi, China in Aug 1957. He received his Ph.D. degree, the M.Sc. and the B.S. degree in Agricultural Engineering at Northwest Agriculture University, Shaanxi, China in 1998, 1984 and 1981respectively. His current major field of study includes intelligent detecting and control, image analysis and recognition and agricultural information technology.

He has been the dean and Ph.D. adviser and professor of the College of Information Engineering, Northwest A&F University, China, in 2003-2009, has been the dean and Ph.D. adviser and professor of the College of Mechanical and Electronic Engineering, Northwest A&F University, China in 2009-2010. He is now working in the College of Mechanical and Electronic Engineering, Northwest A&F University and the managing dean of the "Yangling International Academy of Modern Agriculture", in Yangling, Shaanxi, China. Hitherto, He has published over 120 academic papers in the significant journal, about 30 articles of them has been cited by SCI/EI index. He has also published 2 textbooks as the chief editor and 10 textbooks as the paternal.

Prof. Dongjian He is the member of the Computer Teaching Committee under China's Education Ministry, the chairman of the 9th subcommittee under the 8th degree assessment committee of Northwest A&F University, the vice president both of Shaanxi Image Graphics Society and Chinese Society of Agricultural Engineering, the director both of Shaanxi Computer Teaching Committee and Chinese Society of Image Graphics. Professor He was awarded the honor of Excellent Paper Award by Journal of Chinese Agriculture Engineering in 2002 and 2004, Teaching achievement Award by Education

Department of Shaanxi Province in 2003 and 2004, Excellent Textbook Award by Information Industry.

Linhao Li was born in China, May 23, 1978. He received the M.S. in Computer Science from Stevens Institute of Technology (Hoboken, NJ, USA) in 2012. His research interests are in the area of software engineering, wireless network and services orientated computing.

His current job is a Senior Member of Technical Staff at AT&T Labs Research (Middletown, NJ). He has more than 10 years professional experience adept at large scale data analysis and computational complexity.

Mr. Li is a member of IEEE. His team received AT&T Service Excellence Award (2010) and AT&T Labs President Excellence Award (2011). He won 2nd place of "Shanghai Government Software Design Contest" (2004).

ShaoHua JIANG was born in Hunan China. He received his Ph.D. degree in Spatial Information Science at Huazhong University of Science & Technology, Wuhan, China, and the M.Sc. degree in Computer Science Technology at National University of Defense Technology, Changsha, China. His current major field of study includes Image Processing, Pattern Recognition, Network Security and Software Engineering.

He is working in Hunan Normal University, Changsha, China.

Facial Expression Recognition based on Independent Component Analysis

XiaoHui Guo
China Tobacco GuangXi Industrial CO., LTD

[2]Xiao Zhang, [1]Chao Deng, and [1]Jianyu Wei
[1]China Tobacco GuangXi Industrial CO., LTD
[2]Beihang University, School of Instrumentation Science and Opto-electronics Engineering

Abstract—As an important part of artificial intelligence and pattern recognition, facial expression recognition has drawn much attention recently and numerous methods have been proposed. Feature extraction is the most important part which directly affects the final recognition results. Independent component analysis (ICA) is a subspace analysis method, which is also a novel statistical technique in signal processing and machine learning that aims at finding linear projections of the data that maximize their mutual independence. In this paper, we introduce the basic theory of ICA algorithm in detail and then present the process of facial expression recognition based on ICA model. Finally, we use PCA and ICA algorithm to extract facial features, and then SVM classifier is used for facial expression recognition. Experimental results show ICA is a real effective facial expression recognition method and the recognition rate based on ICA is greater than based on PCA and 2DPCA.

Index Terms—Artificial Intelligence; Pattern Recognition; Facial Expression Recognition; Independent Component Analysis; PCA; SVM

I. INTRODUCTION

Facial expression recognition is short for computer Automatic Facial Expression Recognition is an important part of the artificial psychology theory and computer vision research. It refers to the feature extraction of facial expressions using the computer to be understood, according to human cognition and ways of thinking, classified, and then analyzes the discriminant emotions, such as disgust, surprise, anger, fear, happy, sad and so on. With the rapid development of computer vision technology, facial expression recognition has also been obtained more and more attention, and there are a lot of applications such as human-computer interaction, cognitive science and psychology and so on [1-3].

After decades of continuous research and development, facial expression recognition technology has achieved very important achievements. Face detection, expression feature extraction, facial expression classification and other aspects have a lot of mature and effective algorithms. And they have promoted related disciplines to be rapid development. However, we still see that the field still has a lot of problems to be solved. While humans have a strong ability to identify facial expressions, but the computer achieve quite easy. The main reason is because: First, the human face is a flexible body rather than rigid, hard to facial muscle movement to establish a consistent expression model. Different people different because of age, race, face and facial features, showing the same kind of expression, there will be a considerable difference. This increased the difficulty of expression modeling. Secondly, the expression of the face is a dynamic process, before and after a great deal of relevance in the process. If only extract the expression of a particular time for facial expression recognition, it will lose a lot of information. Finally, expression feature extraction is very difficult because of a wide variety of facial expressions and subtle changes in complex.

Facial expression recognition is an attractive field which is related to image processing, pattern recognition, movement tracking, physiology, psychology and so on. The facial expression recognition system mainly contains three parts: face detection, feature extraction and classification. Among them, feature extraction is the most important part which directly affects the final recognition results [4-5]. Therefore, expression feature extraction is a key step for facial expression recognition. Early facial feature research is mainly based on the geometric features [6]. The basic idea is to use a number of feature points of the person's face, the relative position and the relative distance, and then supplemented with the shape information of the face contour. But its biggest drawback is the recognition accuracy is totally dependent on the extraction of geometric features, these geometric features extraction are very sensitive to changes in illumination, facial expression, gesture, so stability is not high, low recognition rate. Recently, many extraction methods are proposed which are based on statistical characteristics. For example, template matching method [7] which trains these face images in the database as a template, the experimental results show that the performance is significantly superior to the method based on the geometric characteristics.

In face analysis, the dimension of the face image is usually very high, but the distribution of the face image in such a high dimensional space is nosebleed. Therefore, the high dimension is not conducive to the category,

what's more, on the degree of complexity in the calculation is also very and very large. In order to reduce the dimension and extract facial feature, Kirby et al. firstly used the subspace analysis method to realize face recognition [8], this method had been obtained greater success and this subspace analysis method is widely used in recent years. This method then has aroused more and more attention, and thus become one of the mainstream methods of face recognition. The main idea of subspace analysis method is based on certain performance targets to find a linear or non-linear space transform, the original signal data compression to a low-dimensional subspace and then to obtain the more compact distribution of the data in the subspace, which can describe a means of describing data subspace and reduce computational complexity. Recently, these subspace analysis methods used for facial feature extraction including as following: principal component analysis (PCA) [8-9], independent component analysis (ICA) [10], linear discriminant analysis (LDA) [11-12], and non-negative matrix factorization (NMF).

In this paper, we mainly study the facial expression recognition method based on independent component analysis, this is a subspace analysis method. Independent meta-analysis of statistics based on all bands in the solution on principal component analysis and linear discriminant analysis. ICA is as a novel statistical technique in signal processing and machine learning that aims at finding linear projections of the data that maximize their mutual independence. Firstly, we introduce the basic theory of ICA algorithm in detail and then present the process of facial expression recognition based on ICA model. Finally, we use PCA, 2DPCA and ICA algorithm to implement facial expression recognition. Experimental results show ICA is a very effective facial expression recognition method and the recognition rate based on ICA is higher than that of based on PCA and 2DPCA.

II. EXPRESSION FEATURE EXTRACTION

Facial expression recognition is to analysis a particular state for these given expression from facial images and video sequences, and to determine the objects' psychological and emotional state. Facial expression recognition includes the following steps: image acquisition, image preprocessing, face detection, facial feature extraction, classification and recognition. Fig. 1 shows facial expression recognition system diagram.

For the nature of image processing, feature extraction methods can be divided into two categories expression: expression based on a static image feature extraction and dynamic expression video sequences based feature extraction. The former deals with single-frame still face images, the image reflects the general requirements in the expression or greatly exaggerated the state, making the extraction of features more typical expression of such methods include principal component analysis, singular value decomposition and wavelet-based methods. The latter method is used for extracting the characteristic changing process of facial expression.

PCA and 2DPCA, which are also subspace analysis method. PCA is one of widely applied technologies, which has been used in image compression and facial feature extraction. The goal of PCA is to decrease the high dimensional data space into low dimensional feature space. But in face recognition based on PCA, it needs to transform 2D image into 1D image, in order to overcome the flaw, two-dimensional principal component analysis (2DPCA) is proposed to rebuilt image and feature extraction. 2DPCA is based on the image matrix, it does not transform 2D image into 1D image, so it is very simple and more straightforward applied for features extraction. Therefore, we first introduce these two common feature extraction methods.
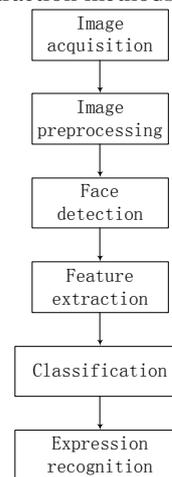

Figure 1. Facial expression recognition system diagram

A. Principal Component Analysis (PCA)

The main idea of PCA is originated from the K-L transform theory, and it is to look for an optimal set of unit which is orthogonal vector-based linear transformation, and with a linear combination of them to reconstruct the original sampling set, to minimize the error between the reconstructed samples and the original samples. PCA is the solution of the eigenvalue problem to obtain the diagonal covariance matrix:

$$S = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T \qquad (1)$$

$$\lambda w = Sw \qquad (2)$$

where N is the total number of samples, $\bar{x}$ represents the mean sample form all sampling set. Calculating the eigenvalues and eigenvectors of the covariance matrix, so we can use the low-dimensional subspace projection coefficients to describe the original data:

$$y = \sum_{i=1}^{t} a_i w_i \qquad (3)$$

where t is the number of the first largest eigenvalues. This feature extraction method based on PCA is proposed by Kirby et al to use for face recognition. Latter, Turk et al. proposed Eigenfaces method for the front of the face recognition. The idea of Eigenfaces is obtained eigenvectors by PCA, the image of any given face can be approximated as a linear combination of the face images of the group characteristics, a combination of the

coefficients will be as face eigenvectors. Original Eigenfaces method is based on Euclidean distance nearest center classifier.

With many applications, there are many recognition methods based on PCA and other subspace methods. Pentland et al. used each perspective for local principal component analysis to achieve the multi-view face recognition. Some researchers used the characteristics of the two main decomposition orthogonal subspace to propose Bayesian framework based on principal component analysis. The selected main priority of Principal component analysis is usually to determine the corresponding eigenvalue according to the size, the eigenvalue is more bigger and it has the higher priority. But on the face recognition, how to choose the number of the main element is the best, the commonly used standard in two ways:

(1) When the corresponding eigenvalue and the largest eigenvalue compared to less than a certain value, to abandon it;

(2) To select eigenvalues and with the total of the characteristic value and the ratio is greater than or equal to 0.9.

Given a $M \times N$ face image, to transform the face image to a column vector firstly, then the dimension of the vector is $D = M \times N$. The mage matrix transforms to vector representation as Fig. 2. Thus, D will be the dimension of the face image, which is also the number of dimension of the image space. Let us assume that there are L vectors of size D representing an image training set, each image is represented as:

$$x_i = [p_1, p_2, p_3, \dots p_N]^T \ i = 1,2, \dots L \qquad (4)$$

The sampling set is mean centered by subtracting the mean image from each image vector. So, mean image μ is represented as followings:

$$\bar{x} = \frac{1}{L}\sum_{i=1}^{L} x_i \qquad (5)$$

Each mean centered image is computed by following equation:

$$\mu_i = x_i - \bar{x} \qquad (6)$$

Then, we need to find a transform matrix which has the largest projection from high dimension data to low dimension onto the vector μ_i. Then to obtain M orthonormal vector e_i, this vector must meet as following:

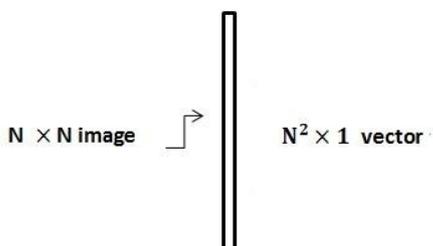$$\lambda_i = \frac{1}{M}\sum_{i=1}^{M}(e_i^T \mu_i)^2 \qquad (7)$$

Figure 2.   Image matrix transforms to vector representation.

These orthonormal vectors are constrained as:

$$e_l^T e_k = \delta_{lk} \qquad (8)$$

In the above equation, e_i and λ_i represent the eigenvectors and the eigenvalues of the covariance matrix respectively, the covariance matrix is computed as:

$$C = WW^T \qquad (9)$$

where $W = \sum \mu_i$. So, the size of the covariance matrix C is $D \times D$, and it could be a high dimension matrix. For example, there is an image and its size is 64×64, then it can create the covariance matrix of size 4096×4096. In solving this question, it is not reasonable to compute the eigenvectors of the covariance matrix directly. A new method is proposed to solve for the eigenvectors of the covariance matrix, we can obtain the eigenvalues and the eigenvectors of the matrix $W^T W$. To assume p_i and α_i represent the eigenvectors and the eigenvalues of the matrix $W^T W$ resprectively, they meet the following equation:

$$W^T W \ p_i \ = \alpha_i p_i \qquad (10)$$

Transform the above equation:

$$WW^T \ (Wp_i) = \alpha_i(Wp_i) \qquad (11)$$

B. 2DPCA

In the face representation technology based on PCA, we need to transform the 2D face image matrices into 1D image vectors. The size of the input image vector is always very big and it is into very high dimension vector space. So, there is very difficult to compute the covariance matrix much accurately, which is due to its big size and the relatively small number of training samples. To overcome the flaws, a new method based on two-dimensional principal component analysis (2DPCA) is proposed for image texture representation and feature extraction. 2DPCA is different to original PCA, which is based on 2D matrices yet PCA is based on 1D vectors. What's more, the original face images do not need to be previously transformed into an image vector. By contrast, 2DPCA can compute under the original image, and the covariance matrix of the training set can be built directly. 2DPCA consists of two advantages over original PCA method:

(1) 2DPCA is easier to evaluate the accuracy of the covariance matrix.

(2) 2DPCA can save some time not to transform 2D image into an image vector. So it has been widely applied rang from many fields.

Given a $M \times N$ face image, we firstly compute the mean image:

$$\bar{X} = \sum_{i=1}^{M} X_i \qquad (12)$$

where X_i represents each face image from sampling set, and it is the 2D image matrix. So, the size of X_i is $M \times N$. Computing the covariance matrix:

$$C = \frac{1}{M}(X_i - \bar{X})(X_i - \bar{X})^T \qquad (13)$$

According to the threshold t, we select the firstly t largest eigenvalues and these related eigenvectors. We note A to be the matrix as the corresponding eigenvectors. The matrix A can be used for feature extraction. For example, an unknown image projection feature is obtained:

$$Y_k = A^T X_k \qquad (14)$$

where Y_k is the projection of X_k. So we can obtain the principal component by using the projection technology. According to the above equation, all projection images are obtained by projection as $Y = [Y_1, Y_2, ... Y_k]$.

III. INDEPENDENT PRINCIPAL COMPONENT ANALYSIS

A. *The Definition of Independent Component Analysis*

Given n observed signals $x_i (i = 1,2, \cdots n)$, assumed each observed signal includes a linear mixed combination with statistically independent source signals $s_j (j = 1,2, \cdots m)$. That is represented as following:

$$X = AS \qquad (15)$$

where $X = [x_1, x_2, \cdots x_n]^T$ represents the observed signal matrix, $S = [s_1, s_2, \cdots s_n]^T$ is source signals matrix, and A is a mixed matrix. The equation (15) represents a basic ICA model, and it describes the observed signals by means to be mixed by the independent component. As we all know, the independent component s_j can't be direct to obtain. What's more, the mixed matrix is also known, and the only known is the observed signals $X = [x_1, x_2, \cdots x_n]^T$. Therefore, the task of ICA is to estimate the mixed matrix A and independent component s_j in the case of the observed signal only to be known.

Because the mixing matrix A is unknown, source signals can't be obtained directly from observed signals. We can obtain a separation matrix W by separating the dependent component from mixed signals, and then we can obtain as following: $\hat{S} = WX$. The equation requires that $\hat{S}$ is a good approximation of the true source signals S. If the separation matrix W is obtained, so the mixed matrix A will be obtained by computing the inverse matrix of W.

Premise without any prior knowledge, these results are certainly not unique to decompose observed signals into multiple mutually statistically independent components. So we want to add some constraints, the results are as close as possible to our expectations. Hyvarine [13] proposed three hypotheses to solve the above question. There hypotheses are as following: Between the respective components of the source signals are statistically independent; the components of the source signal must be non-Gaussian signal can only have a Gaussian signal; the number of components of the source signal is not more than the number of observed signal component, so the mixed matrix is inverse matrix or full column rank mixing matrix, and n observed signal can decompose up to the n source signal components. To satisfy the above three putative ICA model can be estimated. Among them, the first assumption is to solve the problem of independent component analysis, although

this assumption looks very strict, but under normal circumstances would be able to meet. Because most cases, the source signals are sent by different physical systems, generally are able to satisfy the conditions of statistical independence. It is also for this reason, independent component analysis in order to have a wide range in many areas. For the second assumption, we know that the Gaussian distribution is completely symmetrical, and the linear combination of Gaussian distribution is still comply with the Gaussian distribution, so there are two or more components when the source signals obey the Gaussian distribution, the independent component analysis It is impossible to achieve. The third is assumed to be in order to ensure the quantity of the observed signal than the number of source signals, and under normal circumstances, the mixing matrix A is assumed to remain unchanged. In most practical applications, these three are assumed to be able to satisfy.

According to above analysis, there is a certain degree of uncertainty by above method to estimate independent component. Firstly , can't be determined independent component of variance (energy), that use the ICA estimated independent component is only true source best approximation of the signal, and between them there may be some kind of variable proportion, that is independent component may be estimated from its corresponding source signal is multiplied by the weighting coefficients derived results. Second, by the estimated independent component, we can't determine the order in which they are. This is because the S and A are unknown. However, this uncertainty in most cases on the order is irrelevant.

B. *Independent Component Analysis Independence Criterion*

For independent component analysis model, the observed signal consists of a linear mixed signal of a number of independent sources, and the observed signal is closer to a Gaussian distribution, or more the former than the latter Gaussian. So we can put the separation results of non-Gaussian as a metric, which is used to detect the statistical independence between the separation results. When the non-Gaussian for the separation results are very strongest, it indicates that they have the most strongest statistically independence. Entropy is a fundamental concept in information theory, and it is a measure of the signal randomness.

To a random variable x with the probability density function: $p(x)$, the entropy can be defined as:

$$H(x) = -\int p(x) \log p(x) dx \qquad (16)$$

By the information theory of knowledge, we can know all random variables have the same variance, variable Gaussian distribution with maximum entropy. In the base of entropy, the negative entropy can be defined as following:

$$J(x) = H(x_{gauss}) - H(x) \qquad (17)$$

where x_{gauss} and x are with the same covariance Gaussian random variable. From entropy characteristics described above, we can easily deduce that the negative

entropy of the variable is always non-negative. When and only when the variable obeys Gaussian distribution, $J(x) = 0$. The value of $J(x)$ has larger value what indicates the stronger of non-Gaussian quality, so it is more and more far away Gaussian distribution. Therefore, we can also use the negative entropy as an important indicator of the measure of non-Gaussian random variable.

The big advantage of the negative entropy as a measure of non-Gaussian is that it has a rigorous statistical theory background. Therefore, the negative entropy is a measure of non-Gaussian optimal measure. However, we can be seen to calculate negative entropy according to the probability density function of the random variables by definition, which will increase computational complexity. Generally, we use approximated method to calculate the negative entropy.

The second method is mutual information minimization criterion. Mutual information is an important concept in information theory. It is an important indicator of the measure of the independence of the independent component analysis separation results. To a m dimensional random variable $X = [x_1, x_2, \cdots x_m]$, whose mutual information is defined as:

$$I(x) = \sum_{i=1}^{m} H(x_i) - H(x) \qquad (18)$$

where $H(\cdot)$ is the differential entropy. According to knowledge of information theory, $I(x)$ is non-negative. When the random components are statistical independence of each other, $I(x) = 0$. Therefore, mutual information can be used as an important measurement criterion to distinguish signal statistical independence. When the mutual information between signals obtains the minimum value, we can think that these signals are statistically independent.

Mutual information has an important property, so the following equation can be described for reversible linear transformation:

$$I(y_1, y_2, \cdots y_n) = \sum_i H(y_i) - H(X) - log|det\,(W)| \qquad (19)$$

By the definition of the negative entropy, and assume that the components y_i are linear irrelevant each other. Then we can obtain:

$$I(y_1, y_2, \cdots y_n) = C - \sum_i J(y_i) \qquad (20)$$

where C is constant independent of W. The above equation describes the relationship between entropy and mutual information. The negative entropy is obtained the largest value, we consider that these signals are statistically independent. That mutual information reaches the minimum that negative entropy is to obtain maximum value.

The third method is maximum Likelihood criterion. Maximum likelihood estimation is common method to describe independent component analysis. Based on independent component analysis model $X = AS$, The observed signal X is obtained by mixing the source signal S, and the mixed matrix is A. Therefore, the likelihood function of observed signal is defined as:

$$L(A) = E\{log\hat{p}(x)\} = \int p(x)logp(A^{-1}x)dx - log|det\,(A)| \qquad (21)$$

where $\hat{p}(x)$ is the estimated value of the observed signal x with the probability density function $p(x)$. $L(A)$ is a function related with the matrix A. When $W = A^{-1}$, log likelihood function is as following:

$$L(W) \approx \frac{1}{n}\sum_{i=1}^{n}\{logp(Wx)\} + log|det\,(W)| \qquad (22)$$

where n is the number of observed sample which obey independent and identically distributed variables. We can find that maximize the likelihood function to obtain the best estimate of the separation matrix W form. Then we can obtain the best estimate of independent source S. From the perspective of information theory, maximum likelihood estimation criterion essentially mutual information minimization criterion is consistent.

IV. PROPOSED SCHEME

A. ICA Facial Expression Model

According to the aforementioned ICA model, we can obtain:

$$x = As = \sum_{i=1}^{n} a_i s_i \qquad (17)$$

where $x = (x_1, x_2, \cdots x_m)^T$ is observed signals, $s = (s_1, s_2, \cdots s_m)^T$ is the matrix which consists of unknown signal source, and $A = [a_1, a_2, \cdots a_m]$ is $m \times m$ matrix. When meet the blind separation problem under solvable conditions, the separation matrix is obtained to meet $U = Wx$ and U is independent of each other as far as possible.

Each expression image is expanded to become a one-dimensional column vector, and we use X_i to represent the related vector of the i-th image. A training set of M expression images is described as: $X = (X_1, X_2, \cdots X_M)^T$. And assume M vectors consist of M independent basis vectors $S = (S_1, S_2, \cdots S_M)^T$. According to the equation (17), then we know $X = AS$.

Then compute the separation matrix W, so output is as following:

$$U = WX = WAS \qquad (23)$$

$$W = A^{-1} \qquad (24)$$

where $U = (U_1, U_2, \cdots U_M)^T$. U is the estimation of statistical independent base vector, and each row represents a statistically independent basis vectors. The expression vector of the image to be identified is then projected onto the subspace, i.e. the linear combination of the set of base vectors to represent. Let f be identified expression image vector, then:

$$f = a_1 U_1 + a_2 U_2 + \cdots a_M U_M \qquad (25)$$

where $U_1, U_2, \cdots, U_M$ represent M basis vector, and $a_1, a_2, \cdots, a_M$ are projection coefficients. Then to choose SVM classifier to obtain recognition results.

B. SVM Classifier

Support vector machines (Support Vector Machine, SVM) Vapnik et al [14] first proposed in 1995, SVM

classifier is a strong generalization ability, especially in the optimization of the small sample size problem, multi-linear, non-linear and high dimensional pattern recognition problems demonstrate the unique advantage. It can be well applied to the function fitting model predictions other machine learning problem. VC theory SVM method is based on statistical learning theory and based on structural risk minimization principle above, through limited training sample information to seek the best compromise between the model complexity and generalization of learning ability, expect most good generalization ability.

The main idea of SVM is: non-linear sub-sample set, the first sample after the nonlinear transformation projection to a high space to find an optimal classification hyper-plane in high space, making the best classification results. It has a sample data set:

$$(x_1, y_1), \cdots (x_l, y_l), y_i \in \{-1, 1\} \qquad (26)$$

Compute a nonlinear transformation $Z = \Phi(x)$ to the sample data x_l:

$$\begin{cases} w^T z_i + b \geq 1, y_i = 1 \\ w^T z_i + b \leq -1, y_i = -1 \end{cases} \qquad (27)$$

Converse the above equation as following:

$$y_i(w^T z_i + b) \geq 1 \; i = 1, 2, \cdots l \qquad (28)$$

Assume that two kinds of optimal separating surface equation is:

$$w_0^T z + b_0 = 0 \qquad (29)$$

Therefore, the interval between the two categories can be expressed as:

$$\rho(w, b) = min_{\{x|y=1\}} \frac{z^T w}{\|w\|} - max_{\{x|y=-1\}} \frac{z^T w}{\|w\|} \; (30)$$

So this time w_0 should satisfy such that:

$$\rho(w_0, b_0) = \frac{2}{\|w_0\|} = \frac{2}{\|w_0^T w_0\|} \qquad (31)$$

To obtain the maximum value, we use quadratic programming to solve the optimization problem:

$$max_{w,b} \, \Phi(w) = \frac{1}{2}(w^T w) \qquad (32)$$

The formula's constraint condition is $y_i(w^T z_i + b) \geq 1 \; i = 1, 2, \cdots l$. Because the above equation solving linear inseparable problems, then we construct the Lagrangian constraints to solve the minimum question:

$$\Phi(w) = \frac{1}{2} w^T w + \gamma \sum_{i=1}^{l} \xi_i \qquad (33)$$

where ξ classifier error. Then we can obtain:

$$w_0 = \sum_{i=1}^{l} \lambda_i y_i z_i \qquad (34)$$

Decision-making function is as following:

$$f = sgn[\sum_{i=1}^{l} \lambda_i y_i (z^T z_i) + b] \qquad (35)$$

In high-dimensional data transformation kernel function to solve the non-linear data conversion issues,

kernel function method is rewrite the decision function in the above equation to be obtained as following:

$$f = sgn[\sum_{i=1}^{l} \lambda_i y_i K(x_i, x) + b] \qquad (36)$$

We do not need to find a mapping function from low-dimensional to high-dimensional data mapping, only need to know the output can be converted.

For the common linear inseparable, SVM can take advantage of the known nuclear function mapping low-dimensional data from low-dimensional to high-dimensional space, and can be constructed in a high-dimensional space can be divided into a linear hyper-plane. Since the original classic SVM algorithm for the two types of classification and recognition algorithm, achieved by a combination of two types of facial expression recognition of multi-class problems. There are two methods:

(1) "one-to-one" strategy, training multiple classifiers that separate each category twenty-two;

(2) one-to-many "strategy, that is training a classifier which a separate class and all the rest of the class. This paper used the principle of "one to many", and SVM classifier with the nearest neighbor distance separation combined to achieve optimal classification performance

V. EXPERIMENTAL VERIFICATION

In this experiment, we use Jaffe facial expression database, Weizmann database and IMM face database. The database contains 213 images of 7 facial expressions (sadness, disgust, happy, fear, neutral, angry, surprise) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects. Fig. 3 shows the part of Jaffe face database.

Figure 3. Jaffe facial expression database

In order to reduce the position of images of human faces, the impact of the size of the human face, grayscale, size and other factors, the first facial feature points of the image to be normalized, the characteristic parts normalized to the same standards, and then the image further pretreatment, such that the image has the same size, the mean and variance after pretreatment. The normalized characteristic parts criteria: different expression images in the center of the two eyes, the mouth of the center in the same position respectively.

We use PCA and ICA algorithm to implement facial expression recognition. Tab.1 shows expression recognition rate comparison based on PCA and ICA. We can see from the table above, ICA is a real effective facial expression recognition method. The recognition rate is greater than 90%. But the large amount of computation that the ICA spend more computing time.

Next, we select the Weizmann face image database and the IMM face database to verify the recognition results based on ICA algorithm. This database has 28 subjects under 5 different poses, 3 illuminations and 3 facial expressions. We select 3 expressions for each subject to implement experiments. Fig. 4 shows the part of Weizmann database. IMM face database consists of 240 annotated images of 40 different human faces, and each image. Fig. 5 shows the part of IMM face database.

Figure 4.   The part of Weizmann database.

Tab. 1 shows expression recognition rate comparison based on PCA and ICA on Weizmann face database. We can see from the Tab.2, ICA is a real effective facial expression recognition method, and the recognition rate is greater than the recognition method based on PCA.

From Tab. 2, we can see that the recognition rate is biggest using ICA analysis to extract facial expression feature to implement facial expression recognition. In the Jaffe database, we can obtain 86.7% and 88.9% recognition rate by using PCA and 2DPCA respectively. Facial expression recognition can obtain much better performance by using 2DPCA than that of using PCA. Because directly based on two-dimensional sub-image matrix, which can easily reduce the dimension of the original features; in the feature extraction process can completely avoid to use the singular value decomposition method, and it is very simple. According to experimental results from Weizmann and IMM face database, the recognition rate is much higher by using 2DPCA than PCA.

From the experimental results, we also can see that the expression recognition rate is 92.6%, 93.4% and 92.7% by using ICA to extract facial feature from three face databases respectively. Facial expression recognition method based ICA is much better than PCA and 2DPCA. Because the feature extraction method based on ICA consists of principal component analysis and linear discriminant analysis. So, it will have some of their advantages and common characteristics. It can extract facial expression feature more robustness and accuracy, and which will improve the expression recognition rate.

Figure 5.   The part of IMM database

TABLE I.   COMPARISON OF EXPRESSION RECOGNITION RATE

Face database Method	Jaffe	Weizmann	IMM
PCA	86.7%	85.6%	86.8%
2DPCA	88.9%	88.1%	89.0%
ICA	92.6%	93.4%	92.7%

VI.   CONCLUSIONS

Facial expression recognition is one of the most challenging problems in the fields of image processing, biometric identification, movement tracking, computer vision, pattern recognition, physiology, psychology and so on, and it has become a hot research topic in the field of pattern recognition and artificial intelligence recently. Facial expression recognition is all important part of effective computing and intelligent human-machine interactive, which has a wide range of applications and potential market value.

In this paper, we mainly study the facial expression recognition method based on independent component analysis, this is a subspace analysis method. ICA is a novel statistical technique in signal processing and machine learning that aims at finding linear projections of the data that maximize their mutual independence. Firstly, we introduce the basic theory of ICA algorithm in detail, and then present the process of facial expression recognition based on ICA model. Finally, we use PCA, 2DPCA and ICA algorithm to implement expression recognition on different face databases. Experimental results show ICA is a real effective facial expression recognition method and the recognition rate based on ICA is greater than based on PCA and 2DPCA.

REFERENCES

[1]   Bartlett, Marian Stewart, Littlewort Gwen, Fasel Ian, Movellan Javier R, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction", *Conference on Computer Vision and Pattern Recognition Workshop*, pp. 2003.

[2]   Kotsia, I, Pitas, I, "Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and

Support Vector Machines", *IEEE Transactions on Image Processing*, vol.16, pp.172-187, 2007.

[3] Pantic M, Rothkrantz L, "Expert system for automatic analysis of facial expression Image Vision Computing", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.18, no.11, pp. 881-905, 2000.

[4] Ying Zilu, Li Jingwen, Zhang Youwei, "Facial expression recognition based on two dimensional feature extraction", *The 9th International Conference on Signal Processing*, pp.1440-1444, 2008

[5] Lajevardi, S.M, Hussain, Z.M, "Novel higher-order local autocorrelation-like feature extraction methodology for facial expression recognition", *Image Processing*, vol.4, pp.114-119, 2010.

[6] Kyungim Baek , Bruce A. Draper , J. Ross Beveridge , Kai She, "PCA vs. ICA: a comparison on the FERET data set", In Proc. *Of the 4th International Conference on Computer Vision*, 2002

[7] Moghaddam B, pentland A, "Probabilistic visual learning for object representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* pp. 696 - 710, 1997.

[8] M. Turk, A, "Pentland. Eigenfaces for Recognition", *Journal of Cognitive Neuroscience.* vol.3, no.1, pp.71-86, 1991.

[9] Jian Yang, David Zhang, Alejandro F, Jing-yu Yang., "Two-dimension PCA: a new approach to appearance-based face representation and recognition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.26, no.1, 2004.

[10] Marian Stewart Bartlett, Javier R. Movellan, and Terrence J. Sejnowski, "Face recognition by independent Component analysis", *IEEE Trans. on Neural Networks,* vol.13, no.6, 2002.

[11] Juwei Lu, K.N. Plataniotis, and A.N, "Venetsanopoulos. Face Recognition Using LDA Based Algorithms", IEEE *Trans. on Neural Networks,* vol.14, no.1, pp. 117-126, 2002.

[12] Hua Yu, Jie Yang, "A direct LDA algorithm for high-dimensional data – with application to face recognition", *Pattern recognition,* vol.34, pp.2067-2070, 2001.

[13] A. Hyvarinen, E. Oja, "Independent Component Analysis: Algorithms and Applications", Neural Networks, vol.13, pp.411-430, 2000.

[14] Vapnic V, "Statistical learning theory", *New York*: Wiley, 1998.

Total Variation Image De-noising Bases on the Improved Sobel Operator

Jia Fang

College of Sciences, Henan Agricultural University, Zhengzhou, Henan, 450002, China
E-mail: jiafang88@163.com

Cao Qing

College of Sciences, Henan Agricultural University, Zhengzhou, Henan, 450002, China
E-mail: qingcq55@163.com

Abstract—**Traditional Total Variation algorithms often ignore the images edge direction. In order to make up for the flaw of the algorithm existing , a novel denosing method based on direction Total Variation is introduced, on the basis of the combination of the gradient magnitude and orientation. Firstly, the pixels are divided into edge regions and non-edge regions by gradient magnitude. Secondly, the different 4-neighborhood pixels are found in different regions based on gradient orientation. Finally, various of the neighborhoods are analyzed by traditional Total Variation algorithms, so the proposed method can preserves image edge information during image de-noising. Experimental results show that the method combining edge direction information proposed in this paper can improve the neighborhood selection strategy in traditional Total Variation. It not only preserves edges and important details effectively, but also improves the PSNR and visual effect of de-noising images.**

Index Terms—**Edge Orientation, Image De-Noising, TV, Sobel Operator**

I. INTRODUCTION

Image has become a main way to information dissemination. People can acquire a wide range of image data through real life easily. According to some statistics, the information people get from vision is about 80 percent of the total information they get, visible information is intuitionist and reliable. This proverb: one picture is worth a thousand words, clearly reflect graphical unique attributes in information perception. In addition, because of the different imaging pattern and imaging mechanism, the original images contain lots of different types of noise. Noise can severely interrupt people's visual feeling and the mood condition, and decrease image understanding. The extreme condition is so much noise that the image produce distortion, which make storage lost their meaning. It is clear that image denoising is necessary tool to achieve image information. So images de-noising play an important role in improving the image quality.

Images are easy to be polluted by all kinds of noise in its acquisition, transmission or preservation which degrades the quality of original images and has unfavorable influence in the following edge detection of image processing and analysis, feature extraction. Big noise is more likely to cause false rejection and false detection of the structural information, even might have distorted or misunderstood the image content. That is to say, the image de-noising is used to the preprocessing of images before the processing, which can protect the details of the texture information effectively. Traditional de-noising methods is mainly according to the different spectra characteristic between noise and image information. the point-spread function is used to smooth image, i.e. Gaussion smoothing. These de-noising algorithms for a class of smooth regional are very effective. Because noise and edge of image are high frequency signal, it is difficult to distinguish between noise and edge, which make the image details blurred. To achieve much better noise suppression with minimum edge blurring, traditional de-noising processes adopt mathematical morphology. Mathematical morphology is an important element of digital image processing, and it provides for the realization of image processing very powerful method.,which it can effective extract principal features of the signal and repress the noise. The ability of suppress noises bears on the size of structural elements, at the same time single structural element can only have the effect of the geometric feature, other feature and noise are removed together.

According to better energy concentration of image content, sparse signal representation is used to process image with fixed kernels and adaptive kernel. Image can be represented as a linear combination of few fixed function in the fixed kernels. Generally, the image information is mainly distributed in the low frequency domain, whereas noises are in the high frequency part. According to the distribution property of noise and image signal. That is, using a soft threshold to remove noise components from the coefficients of noisy image. The performance of the algorithm is related to threshold selection. Small threshold has very weakly noise suppression capabilities, which can maintain the edge texture characteristic. Big threshold has great interference suppression ability, but Pseudo-Gibbs artifacts is emerged in the edges of the image. To overcome the Pseudo-Gibbs,

BM3D improve sparsity on the basis of the combination of spatial neighborhood and pixel neighborhood. The fixed kernel is built in the case of the noise variance. Since the covariances are not known in practice, all the image content cannot be sparsely represented with a few fixed function. In order to overcome the deficiencies of the fixed kernel, self-adaptive achieve the kernel function of a representational content. Main signal is efficient represented by using these function in the noisy image, and then reached de-noising effects. The ability of restraining noise and preserving edge relate to the selections of the functions, while the selections is only optimal in the sense of linear minimum variance with no regard for the space structure feature.

The noise images have greater pixel values changes than the no-free image. Traditional Total Variation algorithms build a de-noising model adopted 1-norm to measure data approximation based on Image structure information.The algorithm makes an assumption that the pixels locate the edge. Iterations are made by the 4-neighborhood gradient magnitude to image de -noising. This algorithm pays no attention to directional and edge information, which lead to the inability to suppress noise sufficiently in smoothness area, even appear the false edges and the staircase effect. In order to inherit edge-preserving in TV and make up for lack of the algorithm,a novel de-nosing method based on direction Total Variation is introduced, on the basis of the combination of the gradient magnitude and orientation. Firstly, gradient magnitude and direction in image are analyzed based on the optical theory of edge; Secondly, the pixels are divided into edge regions and non-edge regions by gradient magnitude. In the edge regions, the 4-neighborhood pixels nearer the edge are selected based on gradient orientation, while select the space 4-neighborhood pixels in non-edge regions. Finally, the iterative function of Total Variation is analyzed in conjunction with pixel gradient in 4-neighborhood. Experimental results show that the method not only preserves edges and important details effectively, delete the false edges and have good smooth result in smoothness area, but also improves the PSNR and visual effect of de-noising images. Unfortunately the algorithm is not very suitable for weak edge and texture.

II. Noise Classification

Image de-noise is a process of removal noise from the noisy image, based on various kinds of filtering model, such as traditional filtering, wavelet, or partial differential equation and so on. Image de-noising belongs to the image preprocessing stage from the view of image analysis's flow, and belongs to the image restoration from the view of the digital image processing technology. The main performances are as follows:

The de-noising really is the basis of image recognition and analysis. And in general the original images contain lots of different types of noise, which noise can severely interrupt people's visual feeling and the mood condition, and decrease image understanding.

Image de-noising can not only improve the accuracy of vision-based recognition, but also is reliable for further processing. It is necessary to process features extraction, matching method or image fusion and so on in noisy image.

The study of image de-noising can promotes the improvement of performance in other processing steps.

In either image processing system, there are some noise problems in every link that contain the signal acquisition, processing, sending or transmitting and so on, which the image quality can be degraded. The probability of false detection in image segmentation is greater than 0.5% when PSNR is below 14.2dB, and error estimate of parameter is also greater than 0.6%. So study on image de-noising has significance for theories and applications.

According to statistical mechanics, the stationary noise produce probability distributions that are constant in time, whereas it is non-stationary noise. According to statistical of amplitude distribution, probability density functions contain the type of Gaussian, Rayleigh and Impulse. According to the effect of noise on signal or mathematical relationship of image, it can be classified into multiplicative noise and additive noise. Among these, the multiplicative noise includes Gauss noise and salt-and-pepper noise.

Now let assume the input signal $f(x, y)$ to produce a degraded image $g(x, y)$.

Additive noise

Generally speaking, additive noise. is produced by the signal source and shown on output.

$$g(x, y) = f(x, y) + n(x, y) \qquad (1)$$

The image signal add with noise signal, which the feature is $n(x, y)$ is independent of the input signal.

Multiplicative noise

Multiplicative noise has modulating effect on image.

$$g(x, y) = f(x, y) + f(x, y) n(x, y) \qquad (2)$$

where the output is the superposition of these two parts in brief, the second item affected by $f(x, y)$. The more the $f(x, y)$ is, the more the second item is. That is to say, noise is modulated by the image signal, such as the quantum noise, Granulation noise, etc. If the noise and the gray change are small, the second item also is small, which the multiplicative noise can approximately be considered as additive noise.

Through above analysis, may see the vast majority of image noise can be regarded as a normal distribution with mean zero, different standard deviation. the multiplicative noise can approximately be considered as additive noise when the noise and the gray change are small.

3) Pepper-salt noise

The pepper-salt noise, also known as impulse noise, is a model of random white or black point, which is usually due to a camera sensor ,transport channel or decoding, etc. The PDF is given by

$$p(z) = \begin{cases} P_a & z = a \\ P_b & z = b \\ 0 & other \end{cases} \tag{3}$$

where,z denote the gray value. If b>a, gray value b will appear as a light dot in the image. Conversely, level a will appear like a dark dot. If either P_a or P_b is zero, the impulse noise is called unipolar.

III. ORIENTATION EXTRACTION IN THE NEIGHBORHOOD

According to the optical theory of forming the image edge, we can see that the noncontinuous gray grade result in form the edge. In this paper, we can use the Spatial gradient amplitude $|\nabla I(x, y)|$ to describe the pixel value changing speed based on the principle.

$$|\nabla I(x, y)| = \sqrt{\left[\frac{\partial I(x, y)}{\partial x}\right]^2 + \left[\frac{\partial I(x, y)}{\partial y}\right]^2} \tag{4}$$

In order to limit the effects of gradient amplitude (Equation 4)in noise. This article uses Sobel edge-detection operator to compute the gradient. The operator has most smooth filter performance in local area, due to it can restrain the effect of noise to some extent.The template is as following:

$$Sg = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \tag{5}$$

The gradients in horizontal and vertical directions are shown respectively.

$$\begin{cases} \frac{\partial I(x, y)}{\partial x} = Sg * I(x, y) \\ \frac{\partial I(x, y)}{\partial y} = Sg^T * I(x, y) \end{cases} \tag{6}$$

and the edge points is the maximum in local gradient, so adopt the very characteristic of edge to determine whether the pixel point lie on the edge based on the non-maximally suppression principle. The gray value of spatial 4-neighborhood of pixels in non-edge regions has been slower to change, that is to say there are he larger correlation, so the linear representation of the 4-neighborhood pixel is used to show the gray value. These changes are small in the edge. In order to remove noise, at the same time the edge of image can be preserves, so the linear representation of the edge neighborhood pixel is used to show the gray value in the edge. We can use the gradient direction to get neighborhood pixel in the edge based on changing slowly for gray value in the edge(except for corner), as follows

$$\theta(x, y) = \arctan(\frac{\partial I(x, y)}{\partial y} \Big/ \frac{\partial I(x, y)}{\partial x}) \tag{6}$$

In addition, the edge direction is always perpendicular to the direction of the gradient in the 4-neighborhood(the eight-direction)
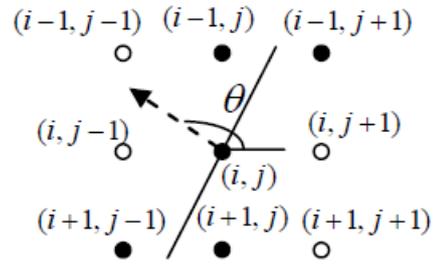

Figure 1.   The edge direction and the neighborhood

Let 0 is placed on the edge, the direction of the point depends on the direction of the gradient, where $S_{i,j}$ denote the orientation neighborhood.

$$S_{i,j} = \begin{cases} \{(i, j-1), (i, j+1)\}, \theta(i, j) = 0 \\ \{(i-1, j-1), (i-1, j), (i+1, j), (i+1, j+1)\}, 0 < \theta(i, j) < \frac{\pi}{4} \\ \{(i-1, j+1), (i+1, j-1)\}, \theta(i, j) = \frac{\pi}{4} \\ \{(i-1, j-1), (i, j-1), (i, j+1), (i+1, j+1)\}, \frac{\pi}{4} < \theta(i, j) < \frac{\pi}{2} \\ \{(i-1, j), (i-1, j)\}, \theta(i, j) = \frac{\pi}{2} \\ \{(i-1, j+1), (i, j+1), (i, j-1), (i+1, j-1)\}, \frac{\pi}{2} < \theta(i, j) < \frac{3\pi}{4} \\ \{(i-1, j-1), (i+1, j+1)\}, \theta(i, j) = \frac{3\pi}{4} \\ \{(i-1, j), (i-1, j+1), (i+1, j+1), (i+1, j)\}, \frac{3\pi}{4} < \theta(i, j) < \pi \end{cases}$$

The neighborhood of the edge pixel need to get pretty close to edge pixel, edge points are smoothed along the the direction of edge. In other words, the method not only reduces noise but also preserves edge information.

IV. DE-NOISES MODEL IN THE ORIENTATION NEIGHBORHOOD

The image de-noising is all very fundamental and an important complicated issues in image analysis understanding. On the other hand, the de-noising comes with the loss of edge information on image details, while the edge is an important parameter which characterizes the image content. Therefore it is vital to maintain more edge information in the process of de-noising. let assume the original clear input signal $I(\mathbf{x})$ to produce a degraded image $I^0(\mathbf{x})$, the model has the following formulation:

$$I^0(\mathbf{x}) = I(\mathbf{x}) + n(\mathbf{x}) \tag{7}$$

where $n(x)$ is a centered Gaussian noise with known variance σ^2 and the noise components are independent. The noise exacerbates the pixel change, which lead to overall gradient changes bigger than the original clear image. In order to remove noise and preserve edges, Traditional Total Variation algorithms build a denoising model adopted 1-norm to measure data approximation based on Image structure information, as follows

$$I(\mathbf{x}) = \arg\min_I \left\{ \iint_\Omega |\nabla I(\mathbf{x})| \, d\Omega \right\} \qquad (8)$$

Integrating the statistical characteristics of the noise, Lagrange multiplier λ is introduced into the objective function to solve the optimization. Redefine a new energy functional:

$$F[I(\mathbf{x}), \lambda] = \iint_\Omega |\nabla_\mathbf{x} I(\mathbf{x})| + \frac{\lambda(I(\mathbf{x}) - I^0(\mathbf{x}))^2}{2} \, d\Omega \quad (9)$$

the a new energy functional is more easily analyzed by using the descent algorithm, we get the following the Eulerian-Lagrangian equation.

$$-\nabla \cdot \left(\frac{\nabla I}{|\nabla I|} \right) + \lambda(I - I^0) = 0 \qquad (10)$$

Based on the analysis of the pixel pixel α and the orientation S, the Equation (9) is converted into iterative calculation

$$I_\alpha = \frac{\sum_{\beta \in S} \left(\frac{1}{|\nabla I|_\beta} + \frac{1}{|\nabla I|_\alpha} \right) I_\beta + \lambda I_\alpha^0}{\lambda + \sum_{\beta \in S} \left(\frac{1}{|\nabla I|_\beta} + \frac{1}{|\nabla I|_\alpha} \right)} \qquad (11)$$

The noisy image contains the detail of the image. The pixel value of the noisy image I_α^0 and orientation neighborhood are adopted in the Equation (10), so it does not cause blurring of the edges of features. Iteration terminates if the Equation (11) is fulfilled, as follow.

$$\sum |I^{(n)}(i,j) - I^{(n-1)}(i,j)| \le 2\% \sum |I^{(n)}(i,j)| \quad (12)$$

This paper has presented Total Variation Image De-noising Bases on the orientation neighborhood, which make edge pixels tend to spread from edge directional and non-edge pixel along with spatial 4-neighborhood. The process is as follows:

The gradient magnitude and directional of the pixel are calculated by Equations (1)-(4).

Adopt the very characteristic of edge to determine whether the pixel point lie on the edge based on the non-maximally suppression principle. if so, the edge directional is vertical direction of the gradient and get orientation neighborhood. Instead, the orientation neighborhood is spatial the 4-neighborhood of pixel.

The diffusion pixel values are calculated by Equations (10).

Repeat steps 1-3 until the termination condition of the algorithm is met.

This processing flow is shown in Figure 2, as follow:

The proposed algorithm use iteration model as an approximation of the original clear image, the pixel value is changed after iteration. So the gradient magnitude and directional of the pixel are recalculated and updated for each iteration. With the increasing of iterations, the orientation neighborhood will be very exactly, so the result is more resemble real-life images.
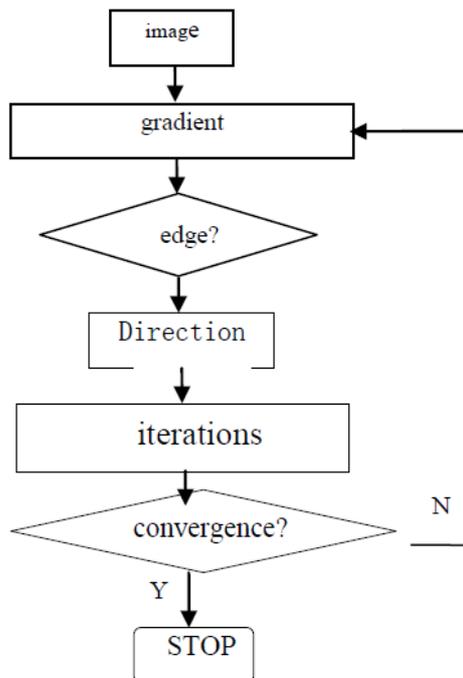

Figure 2.   Algorithm Flow

V.   EXPERIMENTAL ANALYSIS

Traditional Total Variation algorithms make an assumption that the pixels locate the edge. Iterations are made by the 4-neighborhood gradient magnitude to image de-noising. This algorithm pays no attention to directional and edge information, which lead to the inability to suppress noise sufficiently in smoothness area, even appear the false edges and the staircase effect, which cause blurry edge. In order to inherit edge-preserving in TV and make up for lack of the algorithm, a novel de-nosing method based on direction Total Variation is introduced, on the basis of the combination of the gradient magnitude and orientation. By balancing the smoothing area and edge preservation, the control parameter λ is introduced. The Lena image contaminated by an independent Gaussian noise is processed by the different λ, as shown in Figure 3. If $\lambda = 1$, the PSNR is 24.36dB. The fuzzy extent may be slowed by increasing λ. The optimal value of the PSNR is 27.86dB with $\lambda = 2.5$. And as the λ increases, edge information is substantially retained, but the noise is very strong in smoothness region, which cause PSNR loss. If $\lambda = 4.0$, the PSNR is 24.71dB. Known from Equations (10), without considering the structural information when $\lambda = 0$, the proposed algorithm is so smooth that has the problem of edge blurring. Instead, the algorithm make the noise assumed wrongly the structural information is preserved.

Finally, to confirm the algorithm's effectiveness, comparison experiments on the different noise are done using the proposed algorithm ($\lambda = 2.5$), traditional Total Variation algorithms, BM3D, Wavelet algorithm, as the figure 4 shows. The result of the denoise is shown in table 1. From PSNR point of view, the PSNR of the

(a) Noisy image (b) 1.0(24.36dB)

(c) 1.5(25.17dB) (d) 2.0(25.87dB)

e)2.3 (27.37dB) f)2.5 (27.86dB)

g) 2.8 (27.09 dB) h) 3.0 (26.31dB)

i) 4.0(24.71dB)

Figure 3.   Denoising result of difference parameter λ

proposed algorithm is higher than Traditional Total Variation algorithms, BM3D and Wavelet algorithm. In the the visual effects, TV and BM3D have the edge preserving property, while the wavelet algorithm exist the problem of edge blurring of varying degrees. The smooth region has the residual noise using the traditional Total Variation algorithms, while the result of the BM3D is blur, which decreases the contrast of images in some degree. The proposed algorithm reduces the residual noise, cause the reduced edge blurring. The total quality of the processed image is better, the visual effect of the image has been enhanced. But the tiny edge and texture are removed as noise in complex texture, so the excessive smoothing exist in the texture region processed by the proposed algorithm (in Fig. 3-4b)

TABLE I.   PSNR OF DIFFERENT DE-NOISING ALGORITHMS

the original images	The noisy image	De-noising			
		My method	TV	BM3D	Wavelet
Lena	20.19	28.37	27.94	28.53	24.07
	17.10	26.98	26.82	26.65	21.88
	14.34	25.42	23.78	25.81	18.12
	12.56	23.75	21.08	23.45	14.96
	10.69	21.82	20.56	21.21	13.82
peppers	12.95	24.27	22.85	22.38	15.83
	11.53	22.17	21.51	19.75	14.37
	10.75	19.76	17.23	17.37	13.34
	9.16	18.85	17.17	16.61	13.12
	7.64	15.96	14.01	13.19	10.73
baboon	18.01	26.45	26.09	24.78	22.67
	15.12	25.36	24.31	24.02	18.53
	11.56	20.37	20.72	19.17	16.17
	10.15	17.01	16.74	16.21	13.61
	9.15	16.07	14.72	13.87	12.38

VI.   CONCLUSION

Traditional Total Variation algorithms make an assumption that the pixels locate the edge. This algorithm pays no attention to directional and edge information, which lead to the inability to suppress noise sufficiently in smoothness area, even appear the false edges and the staircase effect, which cause blurry edge. In order to inherit edge-preserving in TV and make up for lack of the algorithm,a novel de-nosing method based on direction Total Variation is introduced, on the basis of the combination of the gradient magnitude and orientation. The diffusion of the edge point along with 4-neighborhood is changed in traditional Total Variation algorithms, which make edge pixels tend to spread from edge directional. That is to say, this will ensure maximum smoothness in edge directional and minimum in vertical direction. The result of the experiment validate that the de-noising algorithms based on direction neighborhood Total Variation effectively resolves disadvantages of Traditional Total Variation algorithms, has a faster convergence rate, and realize the compromise with de-noising and edge preserving, greatly improves the objective level and the subjective visual impression.
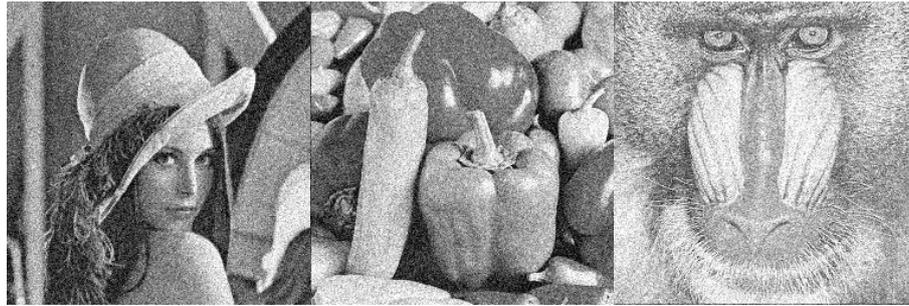
PSNR=16.59dB PSNR=16.76dB PSNR=17.45dB
(a)Noisy image

PSNR=27.34dB PSNR=26.36dB PSNR=26.05dB
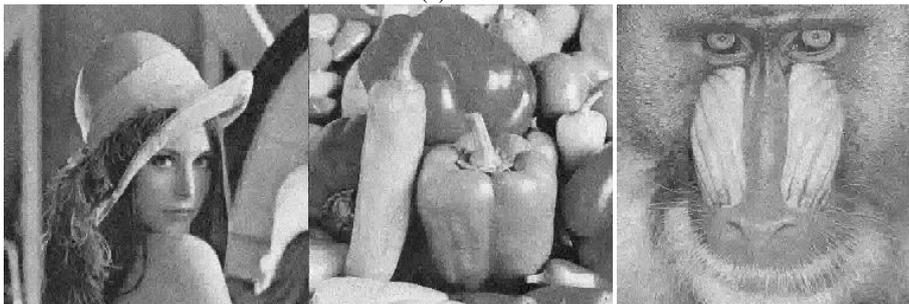(b)the proposed algorithm

PSNR=26.21dB PSNR=24.56dB PSNR=25.05dB
(c)Traditional Total Variation algorithms

PSNR=26.21dB PSNR=24.56dB PSNR=25.05dB
(d)BM3D

PSNR=20.89dB PSNR=20.19dB PSNR=22.03dB
(e)Wavelet algorithm

Figure 4.   Comparisons among different algorithms

REFERENCES

[1]   Katkovnik, Vladimir, Karen Egiazarian, and Jaakko Astola. "Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule." *Journal of Mathematical Imaging and Vision* 16.3 (2002) pp. 223-235.

[2]   Eslami, Ramin, and Hayder Radha. "The contourlet transform for image denoising using cycle spinning." Signals, Systems and Computers, 2003. *Conference Record of the Thirty-Seventh Asilomar Conference on.* Vol. 2. IEEE, 2003.

[3]   Mohideen, S. Kother, S. Arumuga Perumal, and M. Mohamed Sathik. "Image de-noising using discrete wavelet transform." *International Journal of Computer Science and Network Security 8.1* (2008) pp. 213-216.

[4]   Gyaourova, Aglika, Chandrika Kamath, and Imola K. Fodor. "Undecimated wavelet transforms for image de-noising." *Report, Lawrence Livermore National Lab.*, CA 18 (2002).

[5]   Tyler, B. "Interpretation of TOF-SIMS images: multivariate and univariate approaches to image de-noising, image segmentation and compound identification." *Applied surface science 203* (2003) pp. 825-831.

[6]   Dai, Wei, Sheng-lin Yu, and Shuan Sun. "Image de-noising algorithm using adaptive threshold based on contourlet transform." *Acta Electronica Sinica* 35.10 (2007) pp. 1939.

[7]   Hong-juan, Zhang, et al. "A novel image de-noising algorithm combined PCNN with morphology." Intelligent Signal Processing and Communication Systems, 2007. ISPACS 2007. *International Symposium on*. IEEE, 2007.

[8]   Liang, Dong, et al. "A Method for Image De-noising Based on the Contourlet Transform Using Recursive Cycle Spinning." Dianzi Xuebao(Acta Electronica Sinica) 33.11 (2005) pp. 2044-2046.

[9]   Ferguson, Bradley, and Derek Abbott. "De-noising techniques for terahertz responses of biological samples." *Microelectronics Journal* 32.12 (2001) pp. 943-953.

[10] Li, Qingwu, and Chunyuan He. "Application of wavelet threshold to image de-noising." Innovative Computing, Information and Control, 2006. ICICIC'06. *First International Conference on*. Vol. 2. IEEE, 2006.

[11] Wongsawat, Y., K. R. Rao, and S. Oraintara. "Multichannel SVD-based image de-noising." Circuits and Systems, 2005. ISCAS 2005. *IEEE International Symposium on*. IEEE, 2005.

[12] Wang, X. H., R. S. H. Istepanian, and Y. H. Song. "Microarray image de-noising using stationary wavelet transform." Information Technology Applications in Biomedicine, 2003. 4th International *IEEE EMBS Special Topic Conference on.* IEEE, 2003.

[13] Alexander, M. E., et al. "Wavelet domain de-noising of time-courses in MR image sequences." *Magnetic resonance imaging* 18.9 (2000) pp. 1129-1134.

[14] Li, Qingwu, and Chunyuan He. "A new thresholding method in wavelet packet analysis for image de-noising." *Mechatronics and Automation, Proceedings of the 2006 IEEE International Conference on*. IEEE, 2006.

[15] Bruni, Vittoria, and Domenico Vitulano. "Wavelet-based signal de-noising via simple singularities approximation." *Signal Processing* 86.4 (2006): 859-876.

[16] Delakis, Ioannis, Omer Hammad, and Richard I. Kitney. "Wavelet-based de-noising algorithm for images acquired with parallel magnetic resonance imaging (MRI)." *Physics in Medicine and biology* 52.13 (2007): 3741.

[17] Buades, Antoni, Bartomeu Coll, and J-M. Morel. "A non-local algorithm for image denoising." *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 2. IEEE, 2005.

[18] Chang, S. Grace, Bin Yu, and Martin Vetterli. "Adaptive wavelet thresholding for image denoising and compression." *Image Processing, IEEE Transactions on* 9.9 (2000): 1532-1546.

[19] Shih, Yen-Yu, Jyh-Cheng Chen, and Ren-Shyan Liu. "Development of wavelet de-noising technique for PET images." *Computerized Medical Imaging and Graphics* 29.4 (2005): 297-304.

[20] Fleishman, Shachar, Iddo Drori, and Daniel Cohen-Or. "Bilateral mesh denoising." *ACM Transactions on Graphics (TOG).* Vol. 22. No. 3. ACM, 2003.

[21] Deasy, Joseph O., M. Victor Wickerhauser, and Mathieu Picard. "Accelerating Monte Carlo simulations of radiation therapy dose distributions using wavelet threshold de-noising." *Medical physics* 29 (2002): 2366.

[22] Kingsbury, Nick. "Complex wavelets for shift invariant analysis and filtering of signals." *Applied and computational harmonic analysis* 10.3 (2001): 234-253.

[23] Yan, Yanjun, Lisa Ann Osadciw, and Link Hall. "Contourlet based image recovery and de-noising through wireless fading channels." *Proceedings of CISS*. Vol. 5. 2005.

[24] Elad, Michael, and Michal Aharon. "Image denoising via sparse and redundant representations over learned dictionaries." *Image Processing, IEEE Transactions on* 15.12 (2006): 3736-3745.

[25] Dron, J. P., and F. Bolaers. "Improvement of the sensitivity of the scalar indicators (crest factor, kurtosis) using a de-noising method by spectral subtraction: application to the detection of defects in ball bearings." *Journal of Sound and Vibration* 270.1 (2004): 61-73.

[26] Florkowski, M. "Wavelet based partial discharge image de-noising." High Voltage Engineering, 1999. *Eleventh International Symposium on* (Conf. Publ. No. 467). Vol. 5. IET, 1999.

[27] Cancino‐De‐Greiff, Hector F., Rubersy Ramos‐Garcia, and Juan V. Lorenzo‐Ginori. "Signal de‐noising in magnetic resonance spectroscopy using wavelet transforms." *Concepts in Magnetic Resonance* 14.6 (2002): 388-401.

[28] Cancino‐De‐Greiff, Hector F., Rubersy Ramos‐Garcia, and Juan V. Lorenzo‐Ginori. "Signal de‐noising in magnetic resonance spectroscopy using wavelet transforms." *Concepts in Magnetic Resonance* 14.6 (2002): 388-401.

[29] Figueiredo, Mário AT, and Robert D. Nowak. "A bound optimization approach to wavelet-based image deconvolution." *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. Vol. 2. IEEE, 2005.

[30] Elad, Michael, and Michal Aharon. "Image denoising via learned dictionaries and sparse representation." *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.* Vol. 1. IEEE, 2006.

[31] Wang, X. H., Robert SH Istepanian, and Yong Hua Song. "Microarray image enhancement by denoising using stationary wavelet transform." *NanoBioscience, IEEE Transactions on* 2.4 (2003) pp. 184-189.

[32] Katkovnik, Vladimir. "A multiresolution nonparametric regression for spatially adaptive image de-noising." *Signal Processing Letters*, IEEE 11.10 (2004) pp. 798-801.

[33] Bozchalooi, I. Soltani, and Ming Liang. "A smoothness index-guided approach to wavelet parameter selection in

signal de-noising and fault detection." *Journal of Sound and Vibration* 308.1 (2007) pp. 246-267.

[34] Gnanadurai, D., and V. Sadasivam. "Image de-noising using double density wavelet transform based adaptive thresholding technique." *International Journal of Wavelets, Multiresolution and Information Processing* 3.01 (2005) pp. 141-152.

[35] Xiubi, Wang. "Image de-noising based on multi-wavelet." Information Technology and Applications, 2009. IFITA'09. *International Forum on.* Vol. 3. IEEE, 2009.

[36] Hostalkova, Eva, Oldrich Vysata, and Ales Prochazka. "Multi-dimensional biomedical image de-noising using Haar transform." *Digital Signal Processing, 2007 15th International Conference on*. IEEE, 2007.

[37] Lysaker, Marius, Arvid Lundervold, and Xue-Cheng Tai. "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time." *Image Processing, IEEE Transactions on* 12.12 (2003) pp. 1579-1590.

[38] Bilcu, Radu Ciprian, and Markku Vehvilainen. "Combined non-local averaging and intersection of confidence intervals for image de-noising." *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008.

Jia Fang, born in 1976, work at College of Sciences, Henan Agricultural University, Zhengzhou, Henan.

Graphical Analysis of Static and Dynamic Elastic Modulus of Shaving Sheet

Liao Chunhui, Zhang Houjiang, Zhou Lujing, and Yan Haicheng
The School of Technology of Beijing Forestry University Beijing, 100083, China
E-mail: cloudy_c119@163.com.

Abstract—The traditional measurement method of thin wood-based panel's elastic modulus is carried out on the test piece for three-point or four-point bending static deformation test, the determination process is complex, time consuming, and makes damage to specimen. Use materials mechanics quick measuring instrument of sheet board wood, which is self-developed and based on the principle of cantilever's free vibration and cantilever bending to test three standards of thin wood-based panels' and medium density fiberboard's dynamic modulus of elasticity and static bending elastic modulus, use SPSS (Statisti-cal Program for Social Sciences) software to establish the relationship graphics of sheet's dynamic and static bending modulus of elasticity, relationship graphics of dynamic bending and three-point bending elastic modulus, and use linear regression to do correlation analysis on measurement results, while taking advantage of three-point bending to measure same test pieces' three-point bending modulus of elasticity to compare and authenticate. The results show that: the average of cantilever vibration dynamic elastic modulus is higher than cantilever bending static elastic modulus, and there is a very significant linear correlation between them; cantilever vibration dynamic modulus of elasticity and three-point bending modulus of elasticity have significant linear correlation, which shows that taking advantage of materials mechanics fast measuring instrument of sheet wood to measure the modulus of elasticity of thin wood-based panel and medium density fiberboard quickly.

Index Terms—Cantilever Vibration, Cantilever Bending, Dynamic Modulus of Elasticity, Static Bending Modulus of Elasticity, Graphical Analysis

I. INTRODUCTION

Thin wood-based panel refers to the particleboard of which the thickness is less than 8mm, is an important part of the entire particleboard products. It is widely used in furniture manufacturing, automotive industry, construction and decoration industry. Medium density fiberboard is made of small-diameter logs, residues after harvesting, processing and raw materials of non-wood plant fiber, the thickness of 2-30mm, having excellent physical and mechanical properties, decorative properties and processing performance. In order to use Thin wood-based panel rationally and effectively, it should use scientific testing techniques and testing methods to measure its mechanical properties, grasp the characteristics of its mechanical properties, and maximize

the its use value in the most suitable conditions and environment.

The modulus of elasticity (MOE) of thin wood-based panel and medium density fiberboard is one of the important indicators to characterize its mechanical properties. Traditional measurement methods are simply supported static bending, the test process is cumbersome, time-consuming [1-2]. Using vibration method to determine the mechanical properties of wood material, has been proven to be a successful method in the past few decades [3-4]. The measurement results show that there is a very good agreement between the mechanical properties measured by vibration and modulus of elasticity measured by traditional static bending. The most researches of determining mechanical properties of the structure wood based on vibration measurement at home and abroad, are pointing at structure wood of large size, support method of specimen timber are basically simply supported beam support [5-8]. But for sheet of wood materials, due to its light weight, it cannot touch the support well in the process of vibration, so that the vibration signal cannot be sensed well, and thus the method of simply supported beam infeed vibration is infeasible [9].

Based on the cantilever vibration and cantilever bending theory, this article uses self-developed materials mechanics quick measuring instrument of sheet board wood, tests three standards of thin wood-based panels' and medium density fiberboards' dynamic modulus of elasticity and static bending elastic modulus [10]. By the test results, analyzes and discusses the relationship of the dynamic MOE of cantilever vibration and static MOE of cantilever bending. And through universal mechanical testing machine, use three-point bending method to measure the three-point bending modulus of elasticity of shaving sheet [11-12].

Compare and analyze elastic modulus measured by cantilever vibration and bending principle with three-point bending modulus of elasticity, finds that Thin wood-based panel elastic modulus rapid detection method which is based on cantilever free vibration and cantilever bending principle, is feasible [13-15].

II. THEORY

As Figure 1, test piece in cantilever clamping state, deformation occurs after end portion gets loads P. And there is the following relationship:

$$y = \frac{Pl^3}{3E_s I} \qquad (1)$$

where in, P = initial static load (N), y = load point displacement amount (m), L = cantilever overhang (non-clamping) Length (m), I = cross-sectional moment of inertia of cantilever beam (M4), ES = specimen static elastic modulus (Pa).
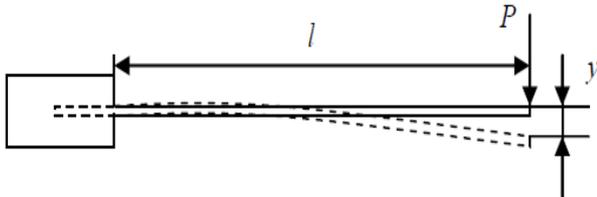

Figure 1. Static bending of a cantilever beam

The formula (1) can be transformed into:

$$E_s = \frac{Pl^3}{3yI} = \frac{4Pl^3}{ybt^3} \qquad (2)$$

where in, b = specimen width (m), t = specimen thickness (m). That is, since the specimen dimensions (l, b, t) are known quantity, after measures the initial static load P and the load point displacement amount y, static elastic modulus Es of specimen can be calculated according to equation (2). This is the basic principle of measuring shaving sheet's static bending elastic modulus based on theory of cantilever bending.

As Figure 2, the first natural circular frequency of free vibration of cantilever beam can be given by the following equation [9]:

$$\omega_{n1} = 2\pi f = \left(\frac{1.875}{l}\right)^2 \sqrt{\frac{E_d I}{m_u}} \qquad (3)$$

where in: ωn1 = the first natural circular frequency (radians / s), f = first natural vibration frequency (Hz), l = cantilever overhang (non-clamping) length (m), Ed = dynamic bending elastic modulus (Pa), I = cantilever cross-sectional moment of inertia (m4), MU = mass of beam per unit length (kg / m).
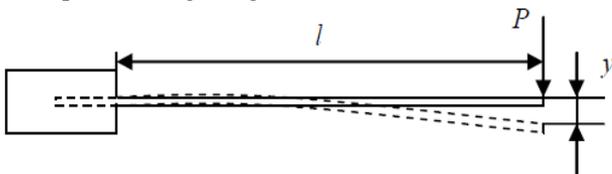

Figure 2. Vibration of a cantilever beam

The formula (3) can be transformed into:

$$E_d = \frac{m_u}{I}(2\pi f)^2 \left(\frac{l}{1.875}\right)^4 = \frac{M}{L}\frac{12}{bt^3}(2\pi f)^2 \left(\frac{l}{1.875}\right)^4 \qquad (4)$$

where in: M = mass (kg) of the test piece, L = total length of the test piece (m), b = specimen width (m), t = thickness of the test piece (m). That is, since the specimen's basic parameters M, L, L, B, t, etc. are known, as long as detects the vibration's first natural frequency f, dynamic modulus of elasticity E_d can be

calculated by formula (4). Formula (4) is an idealized equation, according to flexible vibration theory of Timoshenko, in the condition that the ratio of cantilever length and thickness of the test piece $l / h > 58$ hours, the role of internal shear force and cross-sectional angular movement of the specimen can be ignored. This is the basic principle of measuring shaving sheet's dynamic elastic modulus based on theory of cantilever free vibration.

III. TEST METHODS

A. Test Equipment and Testing Process

Figure 3 is the fast measuring instrument of mechanical properties of sheet wood material, which is developed the research group. The measuring instrument consists of the following components: base, column, specimen holder, laser sensor, adjusting mechanism, initial displacement applying mechanism. where in the specimen holder can adjust the height, play the role of positioning and clamping different lengths of the specimens; initial displacement applying mechanism is installed on the force sensor, which is used to detect the initial load P; further, after the initial displacement applying mechanism releasing the end of the specimen, it can cause free vibration, and sense the vibration signals of the test piece by a laser sensor in the vibration state.

In the test procedure, first adjust the height of the chuck according to specimens with different types and dimension, put the test piece in the specimen clamp. After adjusting the distance between the laser sensor, the force sensor and the test piece according to the experimental requirements, pull the test piece to the retractor, as shown in Figure 4. Running experimental procedures in the computer, first force sensor detects the initial load P signal; then press the retractor to make the specimen vibrate freely, at the moment the laser sensor detects the vibration signal of specimen's end; last the voltage signal produced by force sensor and laser sensor through data acquisition card, the digital signals collected are leaded into the computer to carry out frequency domain analysis and the calculation equation, thereby to obtain the dynamic elastic modulus and the static bending elastic modulus of the test piece.

The three-point bending test which is the contrast agent, is carried out on the Reger RGW-3010-type electronic universal mechanical testing machine. The maximum measured load of the mechanical testing machine can be 10KN, load error is less than ± 1% of the indicated value. Test is in accordance with American Standard ASTM D1037-06a [10].

B. Test Materials

Select three different thicknesses of thin wood-based panels and five different thicknesses of medium density fiberboards as test materials, which are purchased in the market. Depending on the thickness, the thin wood-based panels are processed into specimens of a certain size (thickness × width × length): 3 × 50 × 230mm, 5 × 50 × 340mm, 8 × 50 × 550mm, number of them is 100,100,86. Medium density fiberboards are processed into (thickness

TABLE I.   THE MODULUS TESTED OF FIVE KINDS OF THIN WOOD COMPOSITE

Number	Number of specimen	Code of elastic modulus	Elastic modulus value (GPa)			Standard deviation	Ratio of average MOE (E_s / E_d)
			Average	Min	Max		
PB3×230	100	E_s	2.921	2.108	3.566	0.359	0.91
		E_d	3.214	2.324	3.973	0.39	
		E_b	2.502	1.794	3.206	0.323	
PB5×340	100	E_s	3.027	2.021	3.743	0.631	0.90
		E_d	3.371	2.141	4.265	0.480	
		E_b	2.237	1.203	2.973	0.539	
PB8×550	86	E_s	2.879	2.501	3.465	0.205	0.94
		E_d	3.055	2.611	3.797	0.235	
		E_b	2.423	1.968	3.130	0.214	

TABLE II.   THE MODULUS TESTED OF FIVE KINDS OF THIN WOOD COMPOSITE

Number	Number of specimen	Code of elastic modulus	Elastic modulus value (GPa)			Standard deviation	Ratio of average MOE (E_s / E_d)
			Average	Min	Max		
MDF8.1×550	49	E_s	2.448	1.970	3.112	0.276	0.91
		E_d	2.692	2.146	3.438	0.323	
MDF4.5×340	65	E_s	4.124	3.743	4.743	0.232	0.86
		E_d	4.729	4.214	5.524	0.299	
MDF3.7×340	63	E_s	3.472	3.049	3.865	0.182	0.91
		E_d	3.814	3.400	4.121	0.178	
MDF2.6×340	49	E_s	4.207	3.333	5.583	0.545	0.87
		E_d	4.823	3.952	6.126	0.543	
MDF2.6×230	29	E_s	3.455	2.627	3.900	0.267	0.85
		E_d	4.070	3.141	4.466	0.316	

TABLE III.   THE CORRELATION EQUATION BETWEEN OF E_d AND E_b

Number	Specimen	Linear regression equation: y = ax + b				Correlation coefficient (R)	F value	Sig.	Significance
		y	x	a	b				
PB 3×230	100	E_d	E_s	0.909	-0.0011	0.993	3602.455	0.000	Significant
PB 5×340	100	E_d	E_s	0.7329	0.5572	0.981	1213.992	0.000	Significant
PB 8×550	86	E_d	E_s	0.8496	0.284	0.988	1716.091	0.000	Significant
Total data	286	E_d	E_s	0.7688	0.4691	0.981	3470.029	0.000	Significant

TABLE IV.   THE CORRELATION EQUATION BETWEEN OF E_d AND E_b

Number	Specimen	Linear regression equation: y = ax + b				Correlation coefficient (R)	F value	Sig.	Significance
		y	x	a	b				
PB 3×230	100	E_d	Es	0.727	0.1651	0.858	273. 479	0. 000	Significant
PB 5×340	100	E_d	Es	0.7536	-0.3033	0.882	337. 562	0. 000	Significant
PB 8×550	86	E_d	Es	0.7301	0.1931	0.803	152. 015	0. 000	Significant
Total data	286	E_d	Es	0.6567	0.2708	0.753	368. 37	0. 000	Significant

TABLE V.   THE CORRELATION EQUATION BETWEEN OF E_D AND E_B

| Number	Specimen	Linear regression equation: y = ax + b				Correlation coefficient (R)	F value	Sig.	Significance
		y	x	a	b				
PB 3×230	100	Ed	Eb	0.727	0.1651	0.858	273.479	0.000	Significant
PB 5×340	100	Ed	Eb	0.7536	-0.3033	0.882	337.562	0.000	Significant
PB 8×550	86	Ed	Eb	0.7301	0.1931	0.803	152.015	0.000	Significant
Total data	286	Ed	Eb	0.6567	0.2708	0.753	368.37	0.000	Significant

Figure 3. Cantilever beam tester for thin wood composite
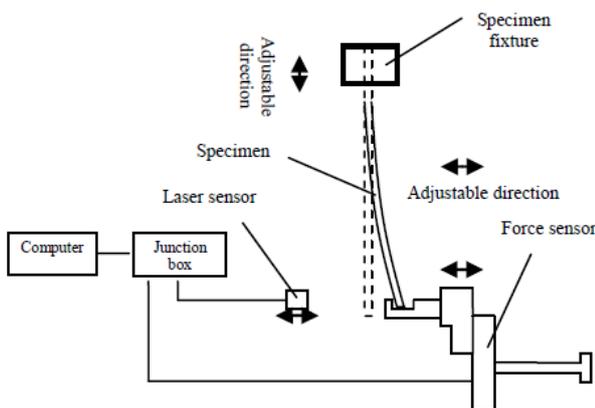
Figure 4. Static bending and vibration of a cantilever beam tes

× width × length): 3.7 × 50 ×340mm, 8.1 × 50 ×550mm, 2.6 × 50 × 230mm, 2.6×50×340mm, 4.5×50×340mm and the number of them is 63, 49, 29, 49, 65. During the test, all clamping length of test pieces is 50mm. Before the test, all specimens are numbered. ID rules are as follows:

(1) PB represents thin wood-based panel, MDF represents Medium density fiberboard;

(2) Numbers following letters represent "thickness × length"(all specimen width of 50mm, it is omitted);

(3) Mantissa represents the serial number of the test pieces.

For example, No. PB3 ×230-10, represent the material quality of test piece is Thin wood-based panel, with dimensions of 3 × 50 × 230 mm (thickness × width × length), is the 10th specimen of the series of 100 specimens.

IV. RESULTS AND ANALYSIS

A. Overall Measurement Results

Specimens of three specifications of thin wood-based panels is 286, specimens of five specifications of medium density fiberboards is 255, the overall measurement results are shown in Table 1 and Table 2, the static elastic modulus E_s measured by cantilever static bending is smaller than dynamic elastic modulus E_d measured by the cantilever vibration. The average modulus of

elasticity indicators of thin wood-based panels E_s / E_d is within a range of 0.90 to 0.94, the average modulus of elasticity indicators of medium density fiberboards E_s / E_d is within a range of 0.85~0.91.

B. Correlation of Dynamic Elastic Modulus E_d and Static Bending Modulus of Elasticity E_s

y=0.909x-0.0011
R =0.993

Figure 5. Relationship of dynamic and static bending modulus of PB 3 ×230

y=0.7329x+0.5572
R =0.981

Figure 6. Relationship of dynamic and static bending modulus of PB 5 ×340

Use self-developed materials mechanics fast measuring instrument of sheet wood to measure the relationship of dynamic elastic modulus E_d and static bending elastic modulus E_s of specimen of three thin wood panel and five kinds of medium density fiberboard, which is shown in Figure 5 to 14. Figure 8 and figure 14 are the comprehensive test data of test pieces of three thin wood panels and five kinds of medium density fiberboards. By multivariate statistical analysis software SPSS (Statisti-Cal Program for Social Sciences), carry out linear regression analysis on test data and comprehensive test data of each sheet specimen, and gets relevant parameters values and correlation coefficient by significant validation of Pearson correlation, which are shown in Table 3 and Table 4.
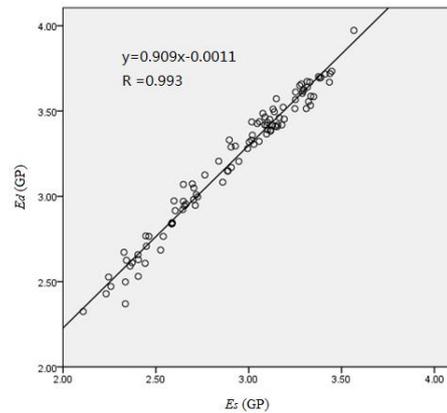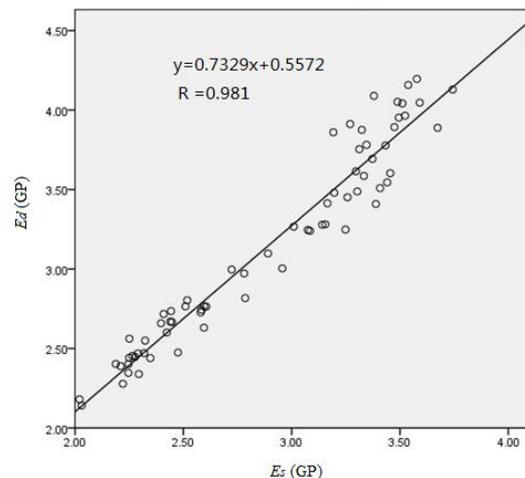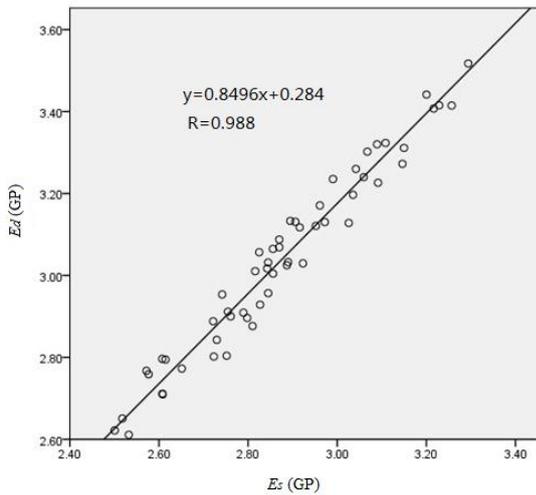
Figure 7.   Relationship of dynamic and static bending modulus of PB 8×550

Figure 8.   Relationship of dynamic and static bending modulus of PB 5 × 340

Figure 9.   Relationship of static and dynamic modulus of MDF 2.6×230

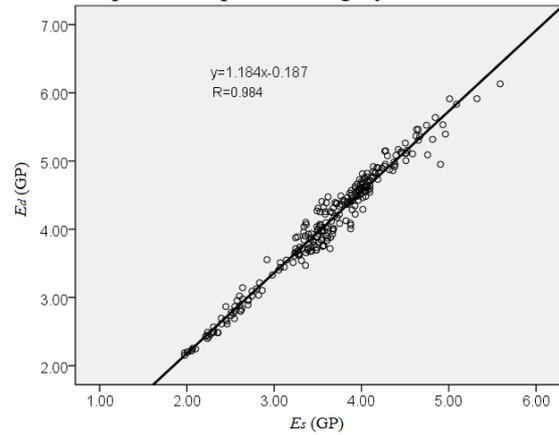and cantilever bending modulus of elasticity of Thin wood-based panel test pieces is highly relevant.

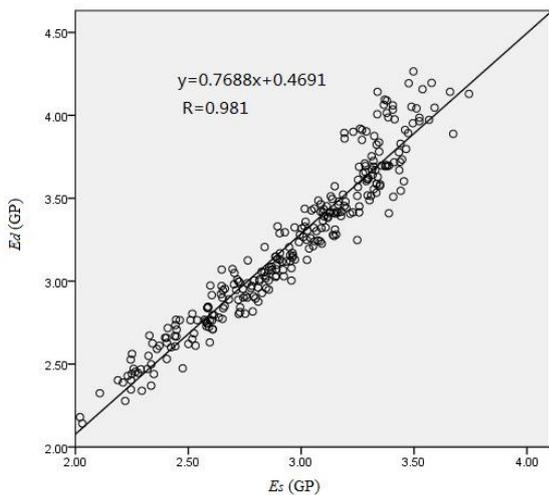Figure 10. Relationship of static and dynamic modulus

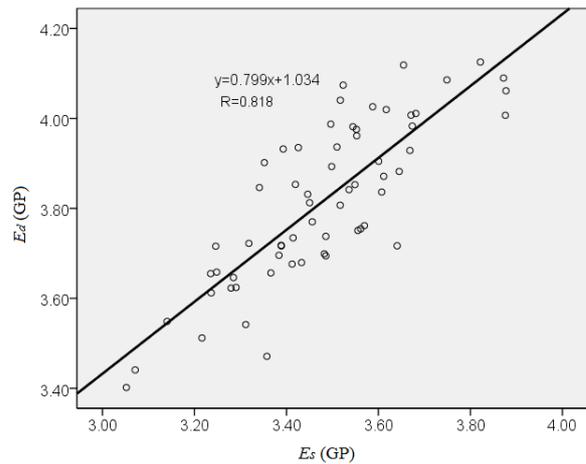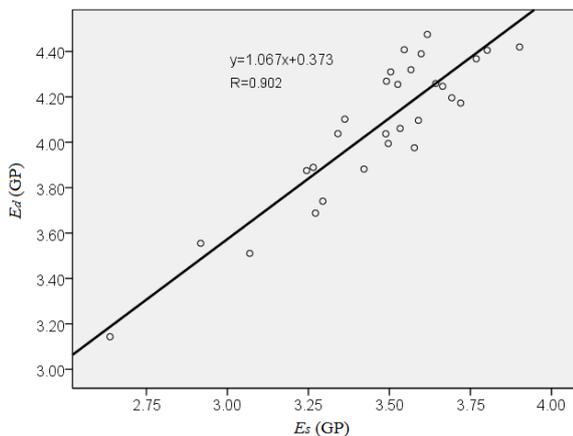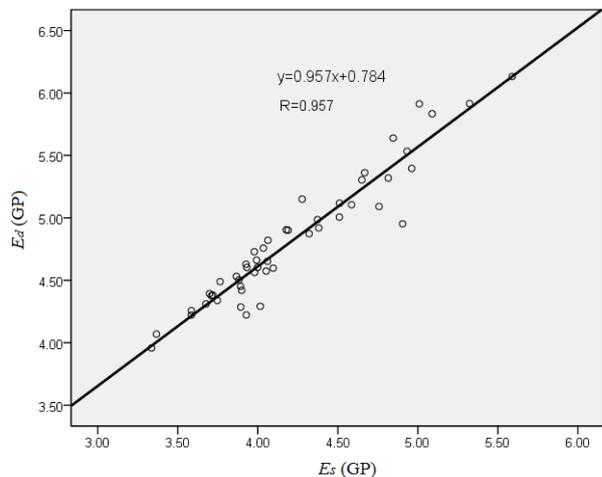Figure 11. Relationship of static and dynamic modulus of MDF 3.7×340

Figure 12. Relationship of static and dynamic modulus of MDF 2.6×340

From Figure 5 to 8 and the relevant data of Table 3, it can be seen that, between Ed and Es of the three shaving sheet and all sheets are significant at the 0.01 level, apparently cantilever vibration dynamic elastic modulus

From Figure 9 to 14 and the relevant data of Table 2, it can be seen that, Ed and Es of the five kinds of sheets and all sheets has a good correlation, correlation coefficients are above 0.8, of which the correlation coefficient of Ed and Es of MDF 8.1 × 550 is the biggest, correlation

coefficient of Ed and Es of MDF 3.7 ×340 is the smallest, is 0.9787 and 0.7595. While five sheet F probability values are greater than 100, indicating that the otherness of cantilever vibration dynamic elastic modulus and cantilever bending modulus of elasticity of test pieces is very significant.
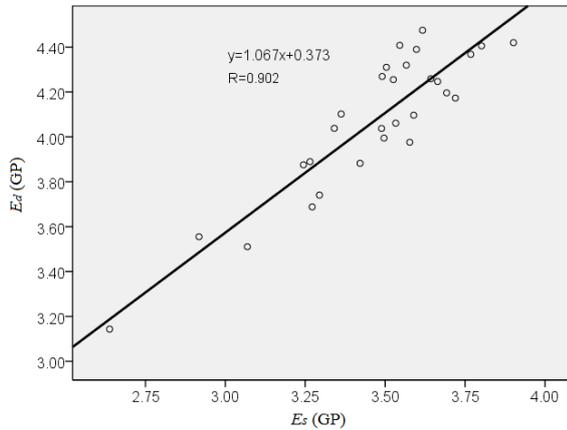

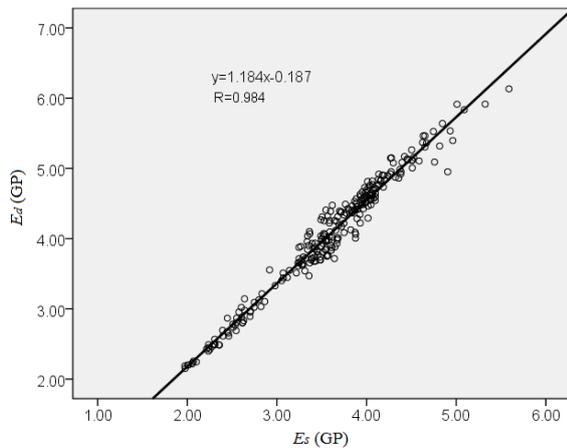Figure 13. Relationship of static and dynamic modulus of MDF 2.6×230

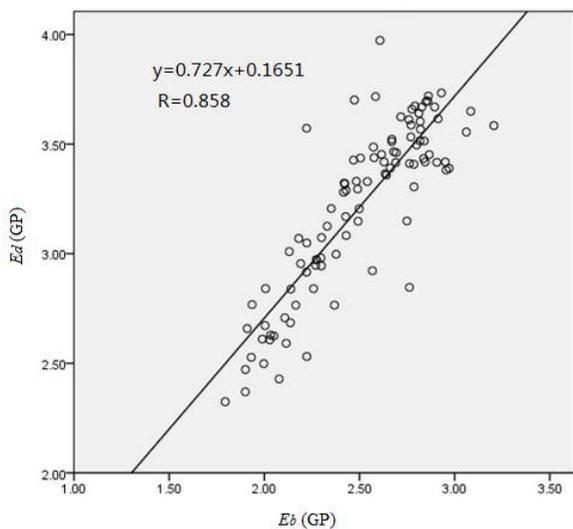Figure 14. Relationship of static and dynamic modulus

Figure 15. Relationship of dynamic and 3-point bending modulus of PB 3 ×230

C. Correlation of Dynamic Modulus of Elasticity Ed and three-point Bending Elasticity Modulus E_b

The relationship of three specimen's dynamic elastic modulus Ed by fast measuring instrument and three-point bending elasticity modulus E_b by universal mechanical testing machine, which is shown in Figure 9 to 12. Figure 12 is the comprehensive test data of three test pieces. By multivariate statistical analysis software SPSS (Statisti-Cal Program for Social Sciences), carry out linear regression analysis on test data and comprehensive test data of each sheet specimen, and gets relevant parameters values and correlation coefficient by significant validation of Pearson correlation, which are shown in Table 3.
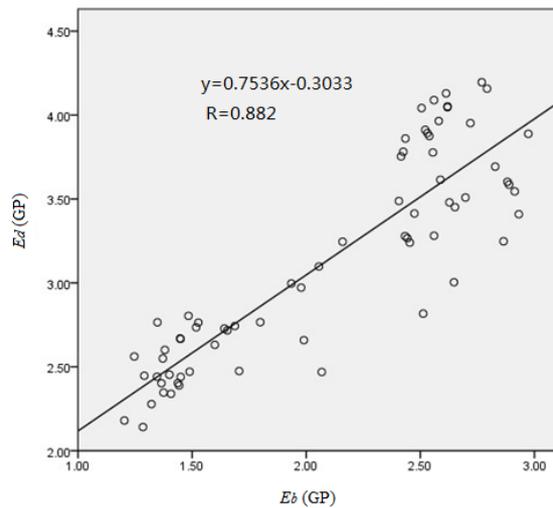

Figure 16. Relationship of dynamic and 3-point bending modulus of PB 5 ×340
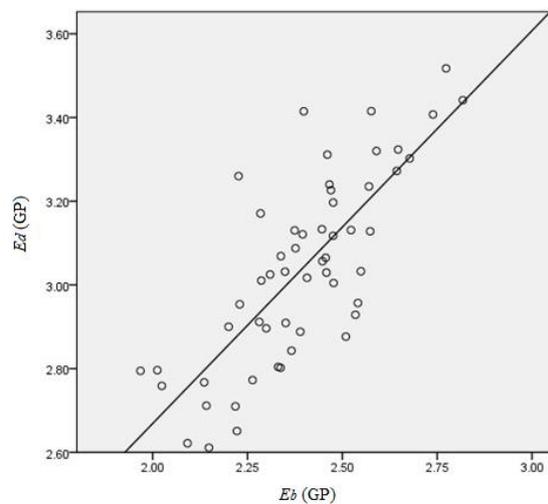
Figure 17. Relationship of dynamic and 3-point bending modulus of PB 8 ×550

From Figure 15 to 18 and the relevant data of Table 5, it can be seen that, between Ed and Eb of the three shaving sheets and all shaving sheets are significant at the 0.01 level, apparently cantilever vibration dynamic elastic modulus and three-point bending modulus of elasticity of Thin wood-based panel test pieces is highly relevant. It indicates that using principle of cantilever vibration to

measure the elastic modulus of thin wood-based panel is feasible.
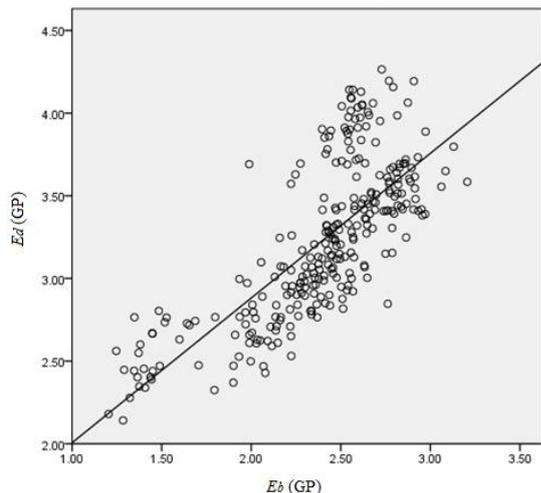

Figure 18. Relationship of dynamic and 3-point bending modulus

V.   CONCLUSIONS

1) As the analysis of experimental data: using the principle of cantilever's free vibration and bending can make rapid measurement of static and dynamic modulus of elasticity of thin wood-based panel and medium density fiberboard.

2) Thin wood-based panel's cantilever static bending modulus value of elasticity is smaller than that of the cantilever vibration dynamic elastic modulus, and are bigger than that of the three-point bending modulus of elasticity; shavings sheet' average ratio of the static bending elasticity modulus and dynamic modulus of elasticity is in the range of 0.90 to 0.94; thin medium density fiberboard's cantilever static bending modulus value of elasticity is smaller than that of the bending vibration dynamic elastic modulus; thin medium density fiberboard's average ratio of the static bending elasticity modulus and dynamic modulus of elasticity is in the range of 0.85 to 0.91.

3) Cantilever vibration dynamic MOE and cantilever static bending MOE of three standards of Thin wood-based panels have significant positive linear correlation, and are significant at the 0.01 level; Cantilever vibration dynamic MOE and cantilever static bending MOE of five standards of thin medium density fiberboards have significant positive linear correlation, and are significant at the 0.01 level.

4) Cantilever vibration dynamic MOE and three-point bending MOE of three standards of thin wood-based panels have significant positive linear correlation, and are all significant at the 0.01 level.

REFERENCES

[1] Zhao Yinjie, Yu Shuiyun. Woodiness Material Science, *China Forestry Publishing House*, pp. 94-99, 2003

[2] Robert J Ross, Brian K Brashaw and Roy F Pellerin. Nondestructive Evaluation of wood, *Forest Products Journal*, Vol. 48, No. 1, pp. 14-19, 1998

[3] Moslemi A A. Dynamic Viscoelasticity in Hardboard, *Forest Products Journal*, Vol. 17 No. 1, pp. 25-33, 1967

[4] Ziqiang Wang, Xia Sun , Manifold Adaptive Kernel Local Fisher Discriminant Analysis for Face Recognition, *Journal of Multimedia*, Vol. 7, No. 6, pp. 387-393, 2012

[5] Jinliang Wan, Yanhui Liu, Multi-Regions Texture Substitution, *Journal of Multimedia*, Vol. 7, No .6, pp. 394-400, 2012

[6] Jing Liu, Guangming Tang, Stego Key Estimation in LSB Steganography, *Journal of Multimedia*, Vol. 7, No. 4, 309-313, 2012

[7] Zhang Houjiang, Shen Shijie, Cui Yingying, Miao Yi. Measuring elasticity modulus of wood using vibration method, *Journal of Beijing Forestry University*, Vol. 27, No. 6, pp. 91-94, 2005

[8] Ross R J, Zerbe J I, Wang Xiping, Green, D W, Pellerin R F. Stress Wave Nondestructive Evaluation of Douglas-Fir Peeler Cores, *Forest Products Journal*, Vol. 55, No. 3, pp. 90-94, 2005

[9] Harris, Cyril M Ed. Harris' Shock and Vibration Handbook, 5th ed. New York: McGraw Hill, 2002.

[10] Shuang Xu, Jifeng Ding, Palmprint Image Processing and Linear Discriminant Analysis Method, *Journal of Multimedia*, Vol. 7, No. 3, pp. 269-276, 2012

[11] Yan Zhao, Hexin Chen, Shigang Wang, Moncef Gabbouj, An Improved Method of Detecting Edge Direction for Spatial Error Concealment, *Journal of Multimedia*, Vol. 7, No. 3, 262-268, 2012

[12] Gang Wang, lin Xiao Gui, DRTEMBB: Dynamic Recommendation Trust Evaluation Model Based on Bidding, *Journal of Multimedia*, Vol. 7, No. 4, pp. 279-288, 2012

[13] Xiaobo Wang, Xianwei Zhou, Junde Song, Hypergraph based Model and Architecture for Planet Surface Networks and Orbit Access, *Journal of Networks*, Vol. 7, No. 4. pp. 723-729, 2012

[14] Jing Ma, Shuli Sun, "Centralized Fusion Estimators for Multi-sensor Systems with Multiplicative Noises and Missing Measurements", *Journal of Networks*, pp. 1538-1545, Vol. 7, No. 10, 2012

[15] Yinpu Zhang, Study on Airspace Covert Communication Algorithm of Covert Communication System, *Journal of Networks*, Vol. 7, No. 4, pp. 730-737, 2012

Chinese Short-Text Classification Based on Topic Model with High-Frequency Feature Expansion

Hu Y. Jun
Sun Yat-sen University / School of Business, Guangzhou, China
Email: hyjsdu96@126.com

Jiang J. Xin and Chang H. You
Sun Yat-sen University / School of Information Science and Technology, Guangzhou, China
Sun Yat-sen University / School of Software, Guangzhou, China
Email: {jx.jiang@foxmail.com, huiyou.chang@hotmail.com}

Abstract—**Short text differs from traditional documents in its shortness and sparseness. Feature extension can ease the problem of high sparseness in the vector space model, but it inevitably introduces noise. To resolve this problem, this paper proposes a high-frequency feature expansion method based on a latent Dirichlet allocation (LDA) topic model. High-frequency features are extracted from each category as the feature space, using LDA to derive latent topics from the corpus, and topic words are extended to the short text. Extensive experiments are conducted on Chinese short messages and news titles. The proposed method for classifying Chinese short texts outperforms conventional classification methods.**

Index Terms—**Chinese short-text classification, feature extension, topic model, high-frequency feature**

I. INTRODUCTION

The emergence of e-commerce, online publishing, instant messaging, and social media has made textual data available in more diverse forms, genres, and formats than ever before. Various data are generated daily: Web news titles by online publishing; queries and questions input by Web search users; Web snippets returned by search engines; blog posts and comments by users on a wide spectrum of online forums, e-communities, and social networks; and online advertising messages from a large number of advertisers.

Many studies have proposed different methods for Web mining and information retrieval (IR) research on textual data. However, few of these methods have been applied to the Chinese context, especially with regard to Chinese short text, which has posed new challenges to Web mining and IR research. This study addresses three main challenges:1) short and sparse data, 2) synonyms and homonyms, and 3) irregular and meaningless words. Classical algorithms, such as latent Dirichlet allocation(LDA)[2], perform well in long text analysis but poorly in short-text classification.

To resolve the sparse feature and noise problems in short-text classification, researchers usually characterize short text by introducing an external corpus or using internal semantics to extract association rules for feature extension. [9] and [15] respectively introduced WordNet and MeSH for short-text feature extension. [1] used Wikipedia as an external corpus with a search engine. Feature extension based on an external corpus is often limited. External corpora are difficult to apply to short texts with highly specialized and specific terminology. The transition point technique (TPT) [12] and self-term extension [13] not only account forward frequency but also extract semantics-associated words based on WordNet ontology to extend features. By contrast, internal semantic association rules unaided by an external corpus establish word association directly from the corpus. [7] extracted association rules from the training corpus and established a high-quality association rule library for feature extension. However, the difficulty was how to choose an appropriate semantic association that improves the performance of short-text classification and reduces the training and learning time overhead. The successful application of the probabilistic topic model in long-text classification has popularized the application of this model in short-text classification. [11] derived a set of hidden topics from a Web corpus through the topic model. [5] used topics on multi-granularity, integrated them, and produced discriminative features for short-text classification.

Motivated by research on the probabilistic topic model, this study proposes a high-frequency feature extension model based on LDA and on principles of word frequency and text classification [4, 12]. Using the LDA topic model to select and expand features can efficiently solve the problem of feature sparseness and noisiness. Extensive experiments were conducted on Chinese short messages and news titles. The proposed method for Chinese short-text classification can significantly improve classification performance compared with traditional algorithms.

The remainder of this paper is organized as follows: Section II reviews related literature, Section III proposesan LDA high-frequency feature (LHF) model, Section IV describes the experimental data, Section V presents the results of applying the proposed model to a

real-world data set and evaluates the methods and results, and section VI concludes.

II. RELATED WORK

A. Semantics

Semantic expansion primarily includes thesaurus and Web knowledge. [16] proposed the use of a search engine, such as Google, to provide more context information. However, the content of short text is difficult to enrich. The method not only consumes more time but also largely depends on the quality of the search engine; the results also tend to be unstable. [10] proposed a three-tier architecture for using Wikipedia and WordNet data to enrich the representation of features. [11] derived topics from an external corpus to enhance the characterization of short texts.[14] directly used the topics of the data set to establish the association of different words in the text. However, this method not only directly models the short text but is also only applicable to distance-based classification and clustering algorithms, such as k-nearest neighbors or k-means.

B. Topic Model

Given the effect of the dimension problem, most statistical models encounter difficulties in learning high-dimensional data. The results of many dimension reduction methods are often characterized by the sparse interpretability of data. The topic model has a wide range of applications. Three primary topic models are in use: latent semantic indexing (LSI) [6], probabilistic LSI (PLSI) [8], and LDA[2]. LSI constructs a text feature vector matrix and then a singular value decomposition of the matrix. PLSI is an LSI model based on the solid statistical theoretical foundation of the probability generation model. LDA is a complete text generation model in which a text is mixed with several implicit or hidden topics mixed with several words. Most researchers directly apply LDA to text classification. Rubin established a statistical topic model to classify multi-label documents and systematically compared the advantages and disadvantages of statistical topic models and discriminative modeling techniques. Feature extension based on topic models has recently received increasing attention. [11] used implicit topics derived from an external corpus to expand text features. [5] proposed a multi-granularity topic model to expand features in short-text classification. [5]modeled multi-granularity topics to obtain an optimal topic for different labels; the model applied to a search-snippet data set outperformed the LDA model(i.e., an increase of 4 percentage points over that of LDA).

III. PROPOSED SCHEME

A. LDA

LDA has broad applications for general discrete datasets, such as text. Specifically, the process of generating a document with n words by LDA can be described as follows [2]:

1. For whole documents, decide on the number of words N according to Poisson distribution $N \mid \xi \sim Poisson(\xi)$.

2. For each document, choose the distribution over topics θ according to Dirichlet distribution $\theta \mid \alpha \sim Dir(\alpha)$.

3. To characterize each of the N words W_n,
 a) Choose a latent topic $Z_n \sim Multinomial(\theta)$.

 b) Choose a word W_n according to multinomial distribution $W_n \mid \{Z_n, \beta_{1:K}\} \sim Multinomial(\beta_{Z_n})$, where α, β, ξ are super parameters. Figure. 1 shows the directed probability graph.
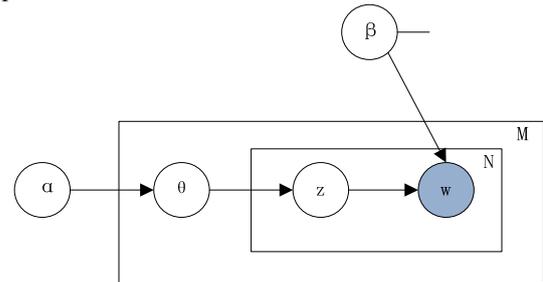

Figure 1. LDA Probability Graph [2]

The above description of creating a new document with n words yields the marginal distribution of the word w_i:

$$p(w_i) = \sum_{j=1}^{K} p(w_i \mid z_i = j)\, p(z_i = j). \quad (1)$$

B. LHF Model

This section presents the LHF model for Chinese short-text classification.

We always use the vector space model to express the text. A text is generally expressed by such feature vectors as $S_d = \{\langle t_1, w_1 \rangle, \langle t_2, w_2 \rangle, \cdots \langle t_{N_d}, w_{N_d} \rangle\}$, where $\langle t_i, w_i \rangle$ is a feature dual group that denotes a feature term w_i with a weight of t_i. This dual group is determined by term frequency–inverse document frequency (TFIDF). Given that the short-text feature is sparse, several short-text feature vectors S_d are null. By contrast, many irrelevant words appear in the short text, contributing nothing to the classification. [4] and [12]studied the effect of word frequency and TPT in short-text classification.

LHF is the high-frequency feature expansion of a topic model. This expansion selects and expands short-text features based on the LDA model. Unlike normal documents, short and sparse documents are usually noisier and less topic-focused. Such documents consist of a dozen words to a few sentences. Thus, whole TFIDF features in documents are noisier; even extracted features selected by CHI can't achieve satisfactory performance due to less topic-focused.

To obtain less noisy and more topic-focused features, this paper proposes feature selection and expansion according to the process of generating a document with n words by LDA for short-text classification. This paper shows such a process with remark labels for each document. The generation scheme is as follows:

TABLE I.
GENERATION FOR LDA

Step 1	For each document $S_d = \{\langle t_1, w_1 \rangle, \langle t_2, w_2 \rangle, \cdots \langle t_{N_d}, w_{N_d} \rangle\}$, $d = 1, 2, \cdots, M$ ·Draw topic proportion $\theta_d \mid \alpha \sim Dir(\alpha)$
Step 2	For each word w_n in that document ·Draw topic assignment $z_n \mid \theta_d \sim \text{Mult}(\theta_d)$ ·Draw word $w_n \mid z_n, \beta_{1:K} \sim \text{Mult}(z_n)$
Step 3	Draw label variable $L_d \mid z_{1:N}, \eta, \sigma^2 \sim N(\eta^\perp \bar{z}, \sigma^2)$ Where $\bar{z} = \left(\frac{1}{N_d}\right) \sum_{n=1}^{N_d} z_n$
Step 4	Then the document can be described as $S_d \propto \{\langle p_1^H, w_1^H \rangle, \langle p_2^H, w_2^H \rangle, \cdots \langle p_{N_{d,h}}^H, w_{N_{d,h}}^H \rangle\}$ $+ \{\langle p_1^L, w_1^L \rangle, \langle p_2^L, w_2^L \rangle, \cdots \langle p_{N_{d,l}}^L, w_{N_{d,l}}^L \rangle\}$ Where $p_n = p(w_n \mid z_n, L_d)$ $N_d = N_{d,h} + N_{d,l}$ $w_n^H \in LHFlib$ $w_n^L \notin LHFlib$

Parameters and variables are shown as follows:
- M: the total number of documents
- N_d: the total number of words in document S_d
- K: the number of latent topics
- α: Dirichlet parameter
- θ_d: topic distribution for document S_d
- $\beta_{1:K}$: topic distribution for word W_n
- $\langle t_i, w_i \rangle$: a feature dual group that denotes a feature term w_i with a weight of t_i; this group was determined by TFIDF
- $\langle p_i, w_i^H \rangle$: a feature dual group that denotes a high-frequency feature of term w_i^H with a weight of p_i; this group is the posterior probability of term w_i^H
- $\langle p_i, w_i^L \rangle$: a feature dual group that denotes a high-frequency feature of term w_i^L with a weight of p_i; this group is the posterior probability of term w_i^L

In LDA, the features of each document can be divided into two parts (Table 1): high-frequency features, which characterize the document, and noise or triviality. This study names the set of all high-frequency features LHFlib, extracts the highest-frequency keyword w_i^H from each category, and obtains new features $\langle p_i, w_i^H \rangle$ through LDA probability generation (Table I). The construction algorithm of LHFlib is as follows:

TABLE II.
CONSTRUCTION OF LDA HIGH-FREQUENCY FEATURE LIBRARY

Step 1	Construct the LDA high-frequency feature library. $LHFlib$
Step 2	Obtain the LDA feature vector of the document: $S_d \propto \{\langle p_1, w_1^H \rangle, \langle p_2, w_2^H \rangle, \cdots \langle p_{N_d}, w_{N_{d,h}}^H \rangle\}$ $+ \{\langle p_1, w_1^L \rangle, \langle p_2, w_2^L \rangle, \cdots \langle p_{N_d}, w_{N_d - N_{d,h}+1}^L \rangle\}$.
Step 3	Extract the high-frequency word vector $\Pi_{M \times N_{d,h}}$ from the library $LHFlib$. $S_d \propto \{\langle p_1, w_1^H \rangle, \langle p_2, w_2^H \rangle, \cdots \langle p_{N_d}, w_{N_{d,h}}^H \rangle\}$, where $p_n = p(w_n \mid z_n, L_d)$ and $w_n^H \in LHFlib$.
Step 4	Obtain the topic distribution feature vector of the document: $\Lambda_{M \times K}$.
Step 5	Merge vectors $\Pi_{M \times N_{d,h}}$ and $\Lambda_{M \times K}$ to obtain LDA high-frequency feature $\Theta_{M \times (N_{d,h}+K)}$.

LHFlib is an LDA high-frequency feature library (Table II). We then extend the high-frequency feature. The algorithm for feature extension for short text is described as follows:

TABLE III.
HIGH-FREQUENCY FEATURE EXTENSION FOR SHORT TEXT

Step 1	For each label L_d: For each document S_d: ·Train the LDA model Φ_k. ·Obtain the top-M words W_d for topic K by posterior Probability.
Step 2	Merge all top-M words for all documents. ·Obtain the LHFlib.

After the LHFs are obtained, we use the support vector machine (SVM) as the classification model. In short-text classification, we not only directly use LHF characteristics but also extend TFIDF features with LHF vector Θ.

A simple flowchart of the LHF method for Chinese short-text classification considering high-frequency feature expansion is shown in Figure 1. The whole process of the LHF model comprises the following:

First, we preprocess the training set to obtain the TFIDF feature vector, and derive an LDA topic model for short texts according to the generation process in Table I.

Second, we construct a high-frequency feature extension library according to Table II.

Third, we extend features to test data according to the algorithm of the high-frequency feature extension above.

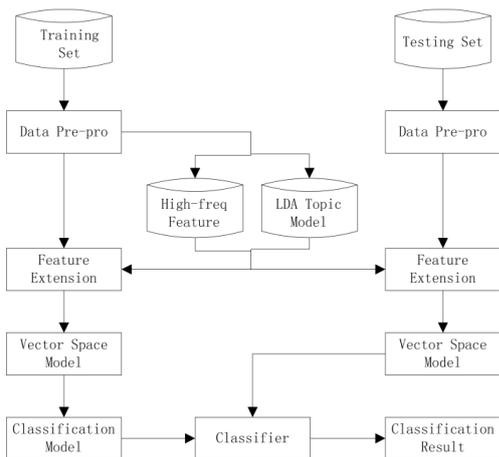Finally, we use the LHF model and SVM to classify the extended short text.

Figure 2. LHF method for short-text classification considering feature extension.

IV. EXPERIMENTAL DATA

We rely mainly on two data sets. First, we use a short message corpus derived from a large short message collection of ABC Telecom. This collection includes more than 20 million messages within a month. Second, we use a news title corpus derived from the Sogou Laboratory. Both collections consist mostly of brief documents, although they have many differences, as described below.

A. Short Message Data

The experiment short message service (SMS) data are preprocessed from the original SMS data set containing more than 20 million short messages, including many fraudulent advertising spam messages. We filter out the spam messages and extract 21 SMS social networks characterized by a transceiver relationship. We use8232 messages from 2551 individuals as our experiment data set. We annotate the sender roles and message topics. The message senders include students, employees, teachers, civil servants, and white-collar workers. Nine message topics are identified: work, family, feeling, learning, leisure, advertising, blessing, and entrusting. The message topics overlap, that is, a message may have another theme. The topic and role distribution of the short messages are shown in Table IV.

TABLE IV.
ROLE AND TOPIC DISTRIBUTION OF SHORT-MESSAGE CORPUS

Topic	Role				
	Student	Worker	Teacher	Government Worker	White-Collar Worker
Work	5	151	298	371	3095
Family	0	65	22	35	375
Love	145	65	128	80	870
Campus	8	0	10	0	29
Leisure	1	0	8	4	121
Advertising	0	0	3	6	10
Blessing	133	19	5	31	455
Social Interaction	14	0	14	4	112
Entertainment	288	0	20	0	308

B. News Title Data

The dataset used in this section is provided by Sogou Laboratory. We extract news titles as our experimental data. A total of 10,953 titles of six categories (culture, society, entertainment, history, military, and reading) are extracted. The length of each short text is 10 to 20 words. We extract the training and testing sets according to the ratio 9: 1. The distribution of the data set is as follows:

TABLE V.
CLASS DISTRIBUTION OF NEWS TITLE CORPUS

No	Class	Training	Testing
1	Culture	1439	160
2	Society	929	104
3	Entertainment	1665	186
4	History	1592	178
5	Military	927	104
6	Reading	3299	367

C. Data Statistics

Figure 3 and Table VI describe the document length distribution of both test data sets. As expected, the documents in the news title corpus are very short (median=6; 9 for SMS), with a compressed distribution of lengths. Most SMS documents are slightly longer than the news titles (median=6), and their lengths vary more significantly than those of the news titles. Most words appear less than 10 times in both data sets (ratio=84.2% in SMS; 93.1% in news title). The corpora analyzed in this study are similar in that the documents are much shorter than documents in standard IR collections. However, the collections also differ from each other.
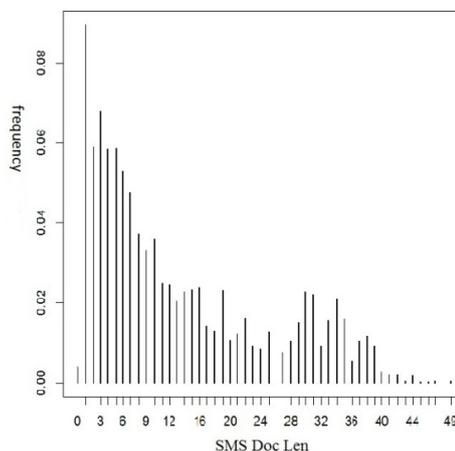

Figure 3. Distribution of Document Length.

TABLE VI.
SUMMARY STATISTICS FOR EXPERIMENTAL CORPORA

Corpus	SMS	News Title
Index Docs	8232	10953
Unique Terms	10282	16109
Unique N&V	3739	8611
Tokens	109868	63875
Median Doc Len	9	6
Mean Doc Len	13.4	5.8
SD Doc Len	130.8	2.8
Ratio	84.2%	93.1%

V. EXPERIMENTAL RESULTS

A. Evaluation Metrics

We use SVM as the classification model. We use the following measures to evaluate the performance of the classifiers: precision(P), recall(R), and F-measure(F1).We also calculate accuracy, macro-P, macro-R, and macro-F1 for macro-assessment:

TABLE VII.
SUMMARY STATISTICS FOR EXPERIMENTAL CORPORA

Precision	TP/(TP+FP)
Recall	TP/(TP+FN)
F1-Measure	$2P \times R/(P+R)$
Accuracy	(TP+TN)/(TP+FP+TN+FN)
Marco-R	$\left(\sum_{i=1}^{k} R_i\right) c_i / N$
Marco-P	$\left(\sum_{i=1}^{k} P_i\right) c_i / N$
Marco-F1	$\left(\sum_{i=1}^{k} F_i\right) c_i / N$

where TP is true positive, FP is false positive, FN is false negative, TN is true negative, N is the total number of documents, and c_i is the total number of documents labeled i.

B. Experimental Results

Table VIII shows the experimental results of the short message data. We use four roles(student, worker, teacher, and civil servants[gov]) as four categories. The two column headings in Table VIII correspond to LDA features alone and expanded features with LHFlib. The proposed LHF model outperforms classical LDA in classifying SMS.

TABLE VIII.
EVALUATION OF TWO METHODS FOR SHORT MESSAGE CLASSIFICATION (%)

Category	LDA Only			LHF Expansion		
	P	R	F1	P	R	F1
Student	77.78	35.00	48.28	78.95	75.00	56.41
Worker	22.22	25.00	23.53	88.89	100.0	61.54
Teacher	45.45	43.48	44.44	86.67	56.52	59.09
Gov	52.50	72.41	60.87	75.68	96.55	70.97
Macro	53.65	49.82	49.11	81.00	80.18	79.19

TABLE IX.
EVALUATION OF TWO METHODS FOR SOGOU NEWS TITLE CLASSIFICATION(%)

Category	LDA Only			LHF Expansion		
	P	R	F1	P	R	F1
Culture	36.36	20.00	25.81	62.04	53.13	57.24
Society	36.00	8.65	13.95	86.30	60.58	71.19
Entertainment	37.67	29.57	33.13	66.67	60.22	63.28
History	51.43	40.45	45.28	75.97	65.73	70.48
Military	53.33	15.38	23.88	75.58	62.50	68.42
Reading	44.63	81.47	57.67	69.85	91.56	79.25
Macro	43.35	44.00	39.56	71.26	70.81	70.14

Table IX shows the experimental result of the Sogou news title data. The data sets include six categories: culture, society, entertainment, history, military, and reading. The two column headings in Table IX similarly

correspond to LDA features alone and to expanded features with LHFlib. The LHF method outperforms classical LDA without feature expansion.

The results show that LHF expansion can obtain less noisy and more topic-focused features. To evaluate the performance of feature selection, we choose the CHI as the base algorithm for the Sogou news title data sets. Figure 4 shows the results.
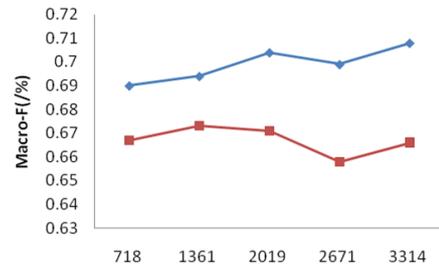

Figure 4.   Evaluation of Two Methods for Short-Text Feature Selection.

LHF exhibits better performance in five different dimensions:781, 1361, 2019, 2671, and 3314, corresponding to 10%, 20%, 30%, 40%, and 50% TFIDF features.

This paper also compares the LHF model with other text classification methods, such as sLDA and maximum entropy modeling (MaxEnt). The LHF model outperforms the other methods in terms of macro-F1.Table X shows the evaluation results of different methods for classifying three short-text corpora. Corpus SMS* has only four roles as in Table VIII and includes 1933 short messages. Corpus Sogou* is extracted from the news titles in Table 5andincludes 3327 news titles. Corpus Yahoo! is obtained from [14] and includes 2400 documents. In the table, M1 refers to LDA; M2,TFIDF+SVM; M3,sLDA[3]; M4,MaxEnt; M5, the short-text classification method in [17]; and M6, the LHF proposed by this paper.

TABLE X.
EVALUATION OF DIFFERENT METHODS FOR CLASSIFICATION OF THREE SHORT-TEXT CORPORA (MACRO-F)

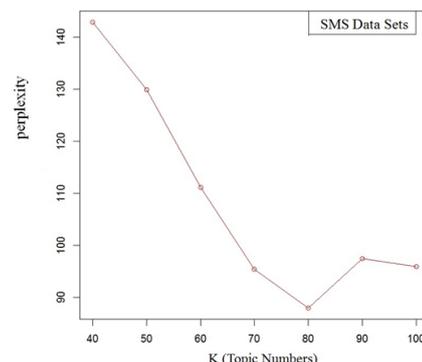| Corpus	M1	M2	M3	M4	M5	M6
SMS	0.491	0.490	0.475	0.480	0.746	0.792
Sogou	0.452	0.367	0.380	0.347	0.708	0.703
Yahoo!	0.333	0.425	NaN	0.397	0.409	0.445

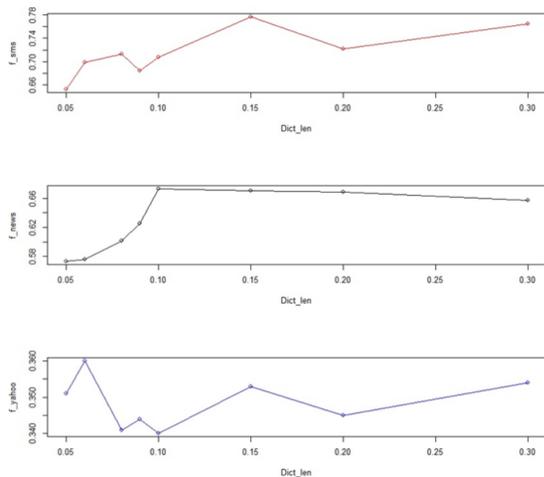Figure 5.   Perplexity indicates the optimal number of latent topics.

Figure 6.   Optimal Length of Library for Three Data Sets.

The optimal number of topics and the length of Lib significantly affect classification performance. This paper used perplexity as the evaluation criterion for the optimal number of topics. The optimal number of topics for the SMS, Sogou, and Yahoo! data sets is 80,80, and 160, respectively. Figure 3 shows the perplexity curve of the SMS data set. $K = 80$ is the optimal number of topics.

Sparseness may induce the optimal length of the LHF library. The optimal library length for SMS, Sogou, and Yahoo! is 15%, 10%, and 6% of the total number of unique terms (Figure 4).

VI. CONCLUSIONS

This study aims to resolve the sparse features and noise problem and thus proposes a short-text classification framework using LDA extended high-frequency features. The experimental result confirmed the effectiveness and feasibility of the algorithm. We draw the following conclusions: (1) The high-frequency words of short texts facilitate the recognition of categories; extracting these words as category features can reduce noise. (2) Introducing high-frequency words through LDA can improve classification performance. When categories have extensive topics, we may have to incorporate other feature extraction methods to improve classification performance.

ACKNOWLEDGMENT

REFERENCES

[1] S. Banerjee, K. Ramanathan, and A. Gupta. Clustering short texts using wikipedia. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2007: 787-788.

[2] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation . the Journal of machine Learning research, 2003, 3: 993-1022.

[3] D.M. Blei, J.D. McAuliffe. Supervised topic models. arXiv preprint arXiv:1003.0783, 2010.

[4] A. Booth. A 'law' of occurrences for words of low frequency. Information and control, 10(4): 386-393, 1967.

[5] Chen M, Jin X, Shen D. Short text classification improved by learning multi-granularity topics. In Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Three. AAAI Press, 2011: 1776-1781.

[6] S. Deerwester. Indexing by latent semantic analysis. Journal of the American Society for Information Science, 1990. 41(6): 391-407.

[7] X. Fan and H. Hu. A new model for chinese short-text classification considering feature extension. In Artificial Intelligence and Computational Intelligence (AICI), 2010 International Conference on. IEEE, 2010, 2: 7-11.

[8] T. Hofmann. Probabilistic latent semantic indexing. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 1999: 50-57.

[9] A. Hotho, S. Staab, and G. Stumme. Ontologies improve text document clustering. In Data Mining, 2003. ICDM 2003. Third IEEE International Conference on. IEEE, 2003: 541-544.

[10] Hu X, Sun N, Zhang C. Exploiting internal and external semantics for the clustering of short texts using world knowledge. In Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009: 919-928.

[11] Phan X H, Nguyen L M, Horiguchi S. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In Proceedings of the 17th international conference on World Wide Web. ACM, 2008: 91-100.

[12] D. Pinto, H. Jiménez-Salazar, and P. Rosso. Clustering abstracts of scientific texts using the transition point technique. Computational Linguistics and Intelligent Text Processing. 2006: 536-546.

[13] D. Pinto, P. Rosso, Y. Benajiba, A. Ahachad, and H. Jiménez-Salazar. Word sense induction in the arabic language: A self-term expansion based approach. In Proc. 7th Conference on Language Engineering of the Egyptian Society of Language Engineering-ESOLE, 2007: 235-245.

[14] X. Quan, G. Liu, Z. Lu, X. Ni, and L. Wenyin. Short text similarity based on probabilistic topics. Knowledge and information systems, 25(3): 473-491, 2010.

[15] D. Reforgiato Recupero. A new unsupervised method for document clustering by using wordnet lexical and conceptual relations. Information Retrieval, 10(6): 563-579, 2007.

[16] Sahami M, Heilman T D. A web-based kernel function for measuring the similarity of short text snippets. In Proceedings of the 15th international conference on World Wide Web. ACM, 2006: 377-386.

[17] Y. Hu, J. Jiang, and H. Chang, "A New Method of Keywords Extraction for Chinese Short-text Classification," New Technology of Library and Information Service, in press.

[18] A.L. Berger, V.J.D. Pietra, S.A.D. Pietra. A maximum entropy approach to natural language processing. Computational linguistics, 1996, 22(1): 39-71.

[19] T.P. Jurka, L. Collingwood, A. Boydstun, et al. RTextTools: Automatic text classification via supervised learning. 2012.

Hu Y. Jun, Guangzhou, October 1978, received the BS degrees in physics from the School of Physics, Shandong University, Jinan, in 2000, and the MS degree in Communication and Information System from School of Information Science and Technology, Sun Yat-sen University, Guangzhou, in 2006. He is now a PhD candidate at School of Business, Sun Yat-sen University. His research interests include natural language processing, text/Web data mining, and complex network analysis.

Jiang J. Xin, Guangzhou, November 1988, received her B.S. degree from South China University of Technology in 2011. She is now a postgraduate student in School of Information Science & Technology at Sun Yat-Sen University of China. Her research interests include natural language processing, machine learning and information retrieval.

Chang H. You, Guangzhou, November 1962, received the BS, MS and PhD degrees in computer science and technology from the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, in 1982, 1985 and 1988, respectively. He is currently a professor of Software College at Sun Yat-Sen University of China. His main research interests include Workflow Management Coalition, Computer Integrated Manufacturing System, Computer Integrated Technology and Application, Artificial Optimization and Intelligent Algorithm, Complex Problem Modeling and algorithm.

Single Image Defogging Algorithm based on Dark Channel Priority

[1,2]Chen Zhen

[1]College of Automation, Harbin Engineering University, Harbin 150001, China
[2]College of Information Technology, Beihua University, Jilin 132013, China
Email: fred61111@sina.com.cn

Shen Jihong
College of Science, Harbin Engineering University, Harbin 150001, China,
Email: shenjihong@hrbeu.edu.cn

Peter Roth
College of Computer Science, Illinois University, Urbana 61801, USA
Email: PeterRoth2042@yahoo.com

Abstract—**Based on dark channel priority, the paper proposes an improved defogging algorithm of single image which can defog the foggy images rapidly. The algorithm in the paper applies the method combining adaptive median filter and bilateral filter to figure out clear dark channel on the edge. And the algorithm is based on the physical model of foggy images to estimate transmission. Compared with the traditional algorithm, the estimated transmission is detailed and clear, and has no need to be optimized, which not only overcomes the disadvantages of traditional algorithm using plenty of time to optimize transmission, but also reduces the complexity of the algorithm. The experimental results indicate that the algorithm realizes rapid and high-quality defogging on single image.**

Index Terms—**Image Restoration, Defogging, Dark Channel Priority, Adaptive Median Filter, Bilateral Filter**

I. INTRODUCTION

Fog is a kind of common natural phenomena. In the misty weather, turbid media of atmosphere such as molecular and suspended particles can create pollution of feedback images and makes fidelity and contrast of the color for images reduce to a great extent. The images as a whole is vague, the details content can't be recognized and the color is biased toward gray and white, which lead to great reduction of dependability for outdoor vision system such as outdoor target recognition, automatic monitor and satellite remote sensing monitoring. Therefore, it has important significance in fog-degraded images clearness.

But fog-degraded images clearness is a challenging task, especially rapid defogging on single image. Because depth information dependent by fog in atmosphere is unknown and acquisition for the depth of single two-dimension image is an unconstrained problem. If multiple images are used to defog, different types of imaging device are needed to shoot multiple images in the same

location [2, 3], which not only has low economic cost, but also is difficult in application for dynamic scene. Or multiple images of the same scene in different weather conditions are collected [4, 5], which is infeasible for outdoor vision system with high real-time demand. Therefore, it is necessary to make rapid defog on single image, which can improve the reliability of all kinds of outdoor vision system in foggy weather.

In recent years, defogging technology for single image has achieved significant progress which is mainly from the following two aspects, one is enhancement algorithm based on image processing, and the other is recovery algorithm based on physical model. As for the former, the typical algorithm is Retinex algorithm which is a kind of model describing color invariant and has a good effect of reinforcing on color image with long-scale contrast formed because of ununiform lighting. Although the image enhanced by Retinex algorithm is on the basis of applying msrcr algorithm with color restoration [6], it is easy for the algorithm to cause color inconsistency or details lost. Because the algorithm needs the user to set multiple gaussian filter scale parameters and the size of parameters and low-frequency color domain of image after enhancement have close relationship with high-frequency details. Recovery algorithm is the the defogging algorithm for single image based on dark channel priority [1] and has better effect. The algorithm can make local repair on the image color in zones according to mist concentration and has good defogging result, and the recovered image color approximates reality and has high definition.

In this paper, an improved haze removal algorithm from a single image is proposed, based on dark channel prior. It can remove haze very fast from the haze image. The dark channel, which is very fine, is calculated with a kind of adaptive median filter and the bilateral filter. The transmission can be estimated, based on the method, from the physical model of imaging in haze. And it is also fine

compared with the transmission estimated by the original algorithm, so that it could be without optimization, and a lot of time is saved, which is spent on refining the transmission by the original algorithm. The complexity of the algorithm is reduced significantly. The experimental results show that the haze image is dehazed very fast with a good quality by the algorithm.

II. PROPOSED SCHEME

A. Defogging Algorithm based on Dark Channel Priority

Dark channel priority is the regularity obtained from the statistics on massive outdoor image without fog. As for most local area except for the sky , there are at least one color channel whose pixel brightness valueapproaches zero in three color channels of GRB. There are three major reasons, the shadow of architectures, people and vehicles, the projection of natural scenes such as trees and rocks, bright-colored objects such as red or yellow flower and leaf, green plant and blue water, and the pixel brightness value of some channels is very low in 3 channels of RGB which also has dark-color surface or objects such as the ground surface and stones. Above all, the scene without fog is full of color, dead colour or shadow, and has dark channels for image. However, brightness valueof dark pixels for images disturbed by fog can be polluted by white light to become higher, that is, the higher the brightness value in ark channels except for the sky, the higher the density of fog in the position. So the dark pixels which are polluted by the fog can be used to estimate concentration of fog and transmission information of light.

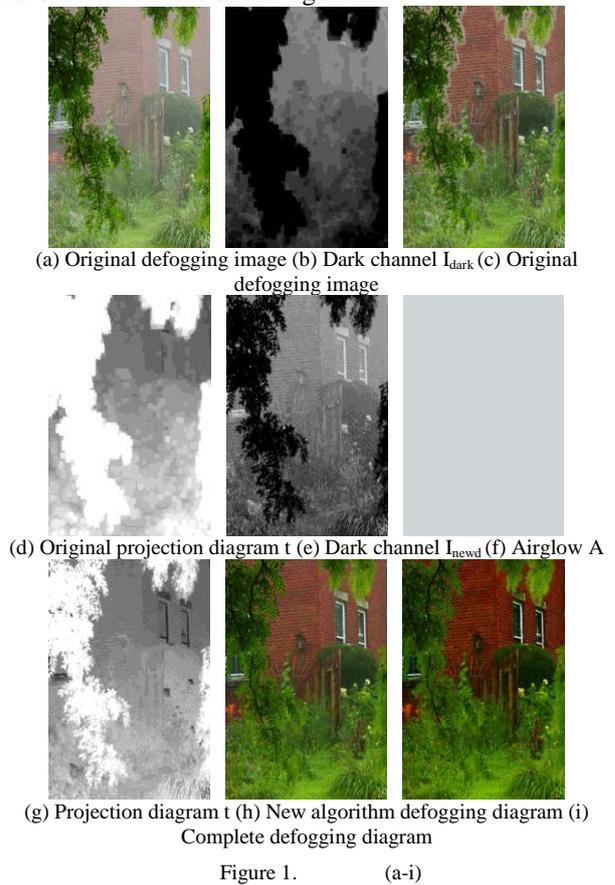With the algorithm in literature [1] which is called H, dark channels can be described :

$$J_{dark}(x) = \min_{c \in \{r,g,b\}} (\min_{y \in \Omega(x)} (J_c(y))) \qquad (1)$$

(1) represents a RGB image without fog, Jc represents a channel in RGB channel of J and Jc is a gray image without fog. $\Omega(x)$ is a piece of square area with x as center and the length of a side takes odd. $\min_{y \in \Omega(x)} (J_c(y))$ is

3 rough gray images with checkerboard effect by taking $\Omega(x)$ as filter to make nonlinear minimum filtering on Jc of three channels. Each pixel of Jdark is the minimum which is taken from three pixels of the corresponding position for 3 gray images. So according to dark channel priority, the intensity of Jdark is always very low and approaches zero as for the area which is not the sky.

I is used to show an atomizing RGB image, dark channels of I obtained according to Formula 1 is Idark to which has rough gray image with checkerboard effect (see Figure 1b), serious halo trace appears at the edge of image after defogging(see Figure 1c). The reason is that transmission picture t is also rough (see Figure 1d). Because HE algorithm uses nonlinear minimum filtering in formula 1 in the process of calculating t, which results in checkerboard effect of t, HE algorithm applies image repair method to optimize the rough transmission diagram t. Although clear image can return after applying

optimized transmission diagram t, the calculation of transmission diagram t is the biggest bottleneck for the algorithm. Side length value of Matting Laplacian giant matrix to be generated in the optimizing process is the area value of defogging images, and each element of Matting Laplacian matrix can be obtained by multiple cycle operation. By testing the image of 800×600 pixel, the whole calculation process takes about 103s, in which optimizing transmission diagram accounts for 94s. The above results are acquired from common PC which has the operating system with Windows XP, CPU with AMD quad core 3.0GHz and memory with 4G RAM. And using 2G RAM can result in memory overflowing, which makes optimization unable to be completed. The algorithm of optimizing transmission diagram is too complicated, which is the biggest obstacle to limit practicability of HE algorithm. However, the paper improves the algorithm which is improved in the paper reduces complexity of calculation significantly in the condition of not influencing the defogging effect, which can shorten runtime of the algorithm.

(a) Original defogging image (b) Dark channel I_dark (c) Original defogging image

(d) Original projection diagram t (e) Dark channel I_newd (f) Airglow A

(g) Projection diagram t (h) New algorithm defogging diagram (i) Complete defogging diagram

Figure 1.   (a-i)

B. Improvement on the Method of Calculating Dark Channels

Firstly, physical model is established for is atomizing image according to the physical property of light transmitting in fog. The following is the description of optical model of atomizing image [12]; the model is commonly used for researching the defogging technology [1, 7, 8].

$$I(x)=J(x)t+A(1-t) \qquad (2)$$

I is RGB image collected in the foggy weather by outdoor visual system and is known. J is RGB image of scenes in the weather without fog, that is, the nearer the atomizing image with J after recovering, the better. t is used to describe the part which has not been scattered in the process of light transmitting visual system for light, that is, the description of the degree of image information in each area is called transmission diagram. A is atmospheric optical components, the brightness of the sky in the weather without fog. Brightness value of each channel for A should be calculated respectively for precision. Obviously, t=1, I(x)=J(x) in the weather without fog, and I is used to estimate t and A to acquire J in the foggy weather.

According to HE algorithm, estimating t and A needs to know Idark, but Idark obtained by HE algorithm is rough, which lead to the roughness of t and spending much time optimizing t, so that the time of calculating defogging is prolonged, and the practicability of the algorithm is limited. The paper proposes a kind of improvement algorithm to calculate dark channels, and the new dark channel is called Inewd and is refined. So t obtained with the improved algorithm is also refined without need to optimize, which shortens the time of the whole process of defogging.

Dark channels of atomization image are estimation for the density of fog. HE algorithm applies the minimum value filtering, which leads to the estimation roughening. But adaptive median filtering in the paper is different from the minimum value filtering, the former approximates to smoothing filtering, and this kind of filtering can retain boundary part in which brightness value of image can change rapidly. The destination is as follows:

kind of
LA: If Zmin<Zmed< Zmax,it turns to LB
or mask size would be increased
If the size is not more than Smax, LA will be repeated.
or Zmed will be output
LB: If Zmin<Zxy< Zmax, Zxy will be output
or Zmed will be output

Mask size is odd and starts from 3. Smax is the biggest mask size. Zxy is the brightness value of coordinate(x,y). Zmin, Zmed and Zmax is respectively the minimum, mid-value and the maximum with (x,y) as center. Next, dark channel Inewd will be calculated, concrete procedures are as follows:

Firstly, the minimum M of each pixel of I in RGB channel is calculated, and M is a gray image:

$$M(x)=\min_{c\in\{r,g,b\}}(I_c(x)) \qquad (3)$$

Next, M uses adaptive median filtering (Smax adaptation takes 1% of the size of atomization image), which can make the value of pixels of dark channels more reasonable at the local range, for example, a brighter local min emerging in the area of uniform brightness can lead to minimal local high light value appearing in M. But using filtering for M can remove hight light value to some extent, and has better edge compared with traditional median filtering.

$$M1=adpmedfilt2(M) \qquad (4)$$

M1 can be as local average of M, and details high-frequency part of M is blurred inevitably. The fuzzy of details generated by median filtering is more slight than the fuzzy of details generated by the minimum value filtering [9]. However, the slight fuzzy also needs to be processed, or detail section for atomization image will be fuzzier after defogging. Bilateral filtering just can dispose the details fuzzy generated with median filtering, which makes the image after defogging is more natural. Bilateral filtering is defined as :

$$BF[I]_p = \frac{1}{W_p}\sum_{q\in S}G_{\sigma_d}(\|p-q\|)G_{\sigma_r}(|I_p-I_q|)I_q \qquad (5)$$

The above is formula 5. Ip and Iq are the brightness value of pixels, p and q. Wp is as follows:

$$W_p = \sum_{q\in S}G_{\sigma_d}(\|p-q\|)G_{\sigma_r}(|I_p-I_q|) \qquad (6)$$

Formula 6 is normalization coefficient. G_{σ_d} and G_{σ_r} is respectively gaussian kernel with σ_d and σ_r as standard deviation on static airspace D and dynamic range R. ‖ represents euclidean distance. So bilateral filtering not only considers spatial relationship of pixels, but also takes similarity relation of brightness value for pixels into consideration. It can smooth images and make images clear [10]. The operation on M1 is as follows:

$$M2=BF(M1,S, \sigma_d ,\sigma_r) \qquad (7)$$

In order to achieve the effect of clearness, S is two times size of adaptive median filter mask, σ_d takes 10 and σ_r takes0.1. BF uses rapid bilateral filtering -in the literatures [14, 15] for improving the operation speed. It firstly calculates one of the biggest dynamic value on the basis of M1 and S, and filters respectively based on size of σ_r , which makes complexity be O(n)

Finally, Inewd should satisfy the constraint condition, 0≤Inewd≤M:

$$Inewd=max(min(M2,M),0) \qquad (8)$$

Dark channel Inewd obtained with the improved algorithm is mist concentration diagram (See Figure 1e). Compared with Idark, the estimating results of the improved algorithm is more meticulous and looks clearer. The pixel whose brightness value is zero in Inewd means the zone without the need to defog. Formula 4, 7 and 8 are merged into function newd for convenient explanation in the next chapter.

C. Estimation of Projection Diagram T and Airglow A

Each pixel of I and J in formula 2 is taken the minimum value in channel RGB

$$\min_{c\in\{r,g,b\}}I_c(x)= t\min_{c\in\{r,g,b\}}J_c(x)+A_c(1-t) \qquad (9)$$

The operation on formula 4, 7 and 8 on both sides of formula 9 can obtain:

$$\text{newd}(\min_{c\in\{r,g,b\}} I_c(x))=\text{newd}(\min_{c\in\{r,g,b\}} J_c(x))t+A_c(1-t) \quad (10)$$

$\text{newd}(\min_{c\in\{r,g,b\}} J_c(x))$ =Jnewd, according to dark channel priority Jnewd=0, $\text{newd}(\min_{c\in\{r,g,b\}} J_c(x))$ =0, and formular 11 is obtained:

$$t=1-\text{newd}(\min_{c\in\{r,g,b\}}\frac{I_c(x)}{A_c}) \quad (11)$$

According to the analysis above, dark channel priority doesn't apply to the sky, but the color of the sky in atomization images is very close to airglow A_c , so $I_c(x)=A_c$, t=0 in the sky. And because the light in the sky is from infinity and transmissivity t is close to zero, formula 11 applies to the sky area and non-sky area, and there is no need to process the sky area individually.

Estimating projection diagram t must estimate airglow A firstly on the basis of formula 11. In the literature [7], pixel value with high-light brightness in image I is regarded as airglow. And most defogging method for single image uses the method to estimate A. However, the brightest pixel in real image may be not in the sky but in a whited object. And s new method of estimating A must be adopted. As mentioned, dark channels of atomization images can obtain rough value of mist density (See Figure 1e). So reliability of estimating airglow can be improved by dark channels. Firstly, the pixel of 0.1% with the biggest brightness value in dark channels can be taken. And the corresponding positions of these pixels in I can be found, then the pixels responding to the positions can be found in three channels of I respectively. Finally, the biggest value of those pixels is taken from 3 channels of I to be as the brightness value of all pixels for the channels responding to A. And A can be estimated (See Figure 1f), and we should notice that A is not the brightest point in image I. The method estimating airglow A based on dark channels is more reliable than the method in literature [7].

Transmission diagram can be estimated through formula 11. But in reality, the air contains some impurity inevitably in the condition without fog. Therefore, the fog still exists for remote objects. It is the existence of fog that people feels the existence of depth, which is called areal perspective [11]. If the fog is removed completely, images look not true and the depth disappear (See Figure 1i). Therefore, a constant ω (0<ω≤ 1) can be set in formula 11 and the fog covered on the remote objects is reserved, and the modification makes defogging diagram more perfect. The value of ω is set according to concrete conditions. In the paper, ω selects a constant value, 0.85.

$I_c(x)$ / A_c =N, $\text{newd}(\min_{c\in\{r,g,b\}}\frac{I_c(x)}{A_c})$ is using the improved algorithm to acquire dark channel of N and is set as N_{newd} , transmission diagram t (See Figure 1g) is

$$t=1-0.85N_{newd} \quad (12)$$

D. Acquisition of Images after Defogging

According to A and t estimated in 2.2 and the formula 2, we can obtain:

$$J(x)=\frac{I(x)-A}{t}+A \quad (13)$$

Transmissivity can't be zero, so t is set as a lower limit t0. The paper takes t0=0.1, and the final defogging image j (See Figure 1h) is:

$$J(x)=\frac{I(x)-A}{\max(t,0.1)}+A \quad (14)$$

III. SIMULATION ANALYSIS

In order to verify the practicability and effectiveness of the improved algorithm, Matlab 2012a is used to make simulation experiment on common pc which has the operating system with Windows XP, CPU with AMD quad core 3.0GHz and memory with 4G RAM. The images of cities, outskirts and remote sensing which are polluted by fog are defogged. HE algorithm based on msrcr of retinex is applied. Visual effect, operation speed and objective image resolution are compared. The simulation results are shown in Figure 2, Figure 3 and Figure 4.
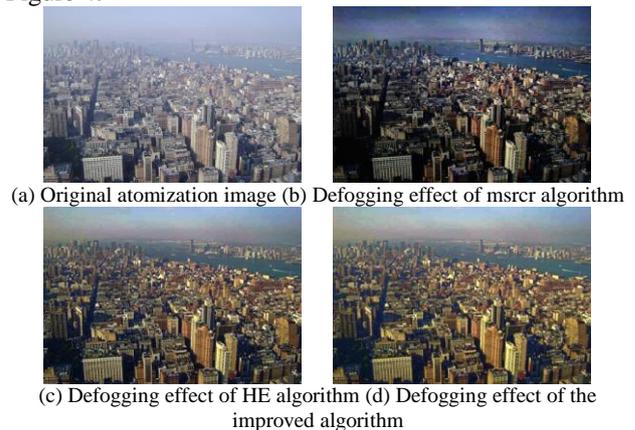

(a) Original atomization image (b) Defogging effect of msrcr algorithm

(c) Defogging effect of HE algorithm (d) Defogging effect of the improved algorithm

Figure 2.   (a-b) Comparison of simulation effect for city images

(a) Original atomization image (b) Defogging effect of msrcr algorithm

(c) Defogging effect of HE algorithm (d) Defogging effect of the improved algorithm
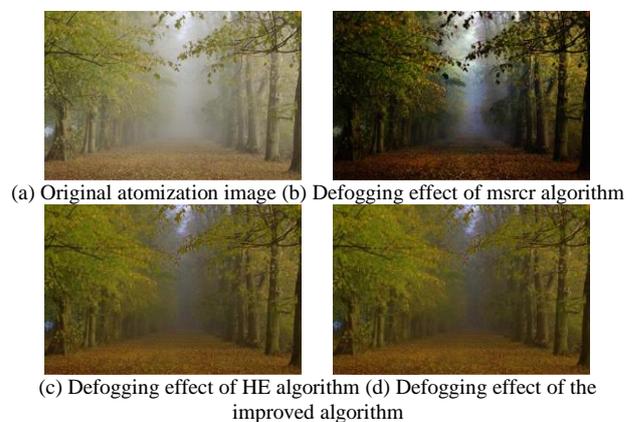
Figure 3.   (a-b) Comparison of simulation effect for outskirts image

Three algorithms realizes the defogging to some extent. As mentioned, the color for defogging images with msrcr algorithm are somewhat distorted, for example, the color in the sky of Figure 2b bias dark and the transition is not natural, and the color of leaf in Figure 3b and the color of ground surface in Figure 4b. He algorithm and the improved algorithm all achieve better effect of defogging. For example, the color of the sky in Figure 2d is natural and the level is clear, the defogging effect of the woods in Figure 3d is good, and the texture of ground surface in Figure 4d is more clear. Most of all, the new transmission diagram t is very clear. There is no halo formation left after defogging on the leaf in Figure 3a (See Figure 3d), and the detail texture of leaf in left side of Figure 3d is more clear because of the application of bilateral filtering.

(a) Original atomization image (b) Defogging effect of msrcr algorithm

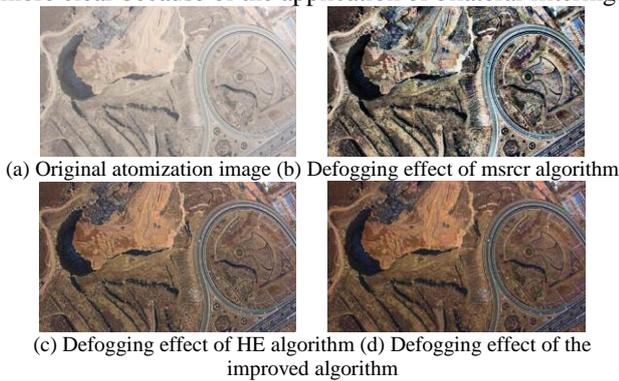(c) Defogging effect of HE algorithm (d) Defogging effect of the improved algorithm

Figure 4.   (a-b) Comparison of effect for remote sensing images

Next, various algorithms are compared objectively. And histogram, mean value, standard deviation and entropy of R channel from image 4(a) to (d) are calculated, and the results are shown in Figure 5 and Table 1.

TABLE I.   MEAN VALUE AND STANDARD DEVIATION OF R CHANNEL FROM IMAGE 4(A) TO (D)

|	Mean value	Standard deviation
Figure 4(a)	183.5311	32.9575
Figure 4(b)	118.5674	65.2248
Figure 4(c)	116.6221	42.7048
Figure 4(d)	109.1401	45.7910

From Figure 5, we can see that the pixel stretch of histogram for the algorithm in the paper is greater and has evident spike, which indicates that the image is clear and there is high contrast. And we can see from Table 1 that the mean value of algorithm in the paper is the minimum, which shows that the interference of fog is the minimal. But it has an appropriate standard deviation, which indicates that the image includes the most information and is the clearest.

Then statistical evaluation method of image quality which is proposed in literature [22] was applied. And the statistical results show that when gray average is between 100 and 200, and the standard deviation is between 35 and 80, the quality of the images is better, as shown in Figure 6. And concrete steps of statistics are: the image is decomposed into different subblocks (50×50), and the standard deviation σ of each subblock is figured out and is for averaging, and multiplying the average standard
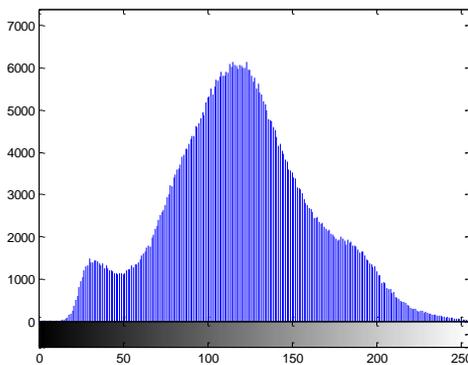
deviation σ and gray average of image can get evaluation result. The greater the multiplying result is, the better the quality of image is. The statistical results of image are shown in Table 2.

(a)

(b)

(c)
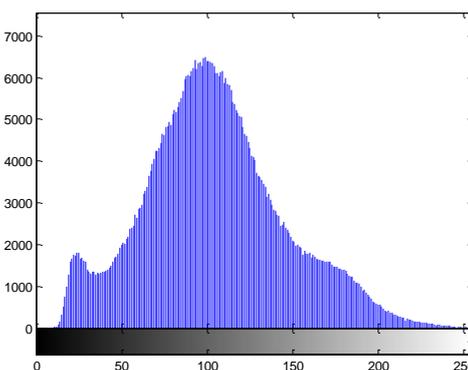
(d)
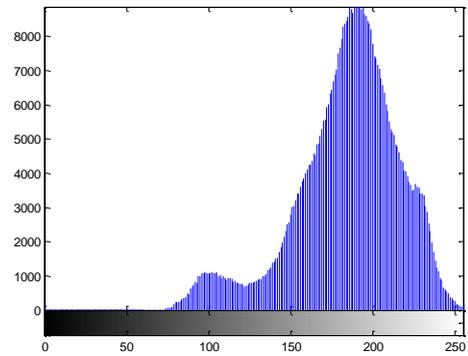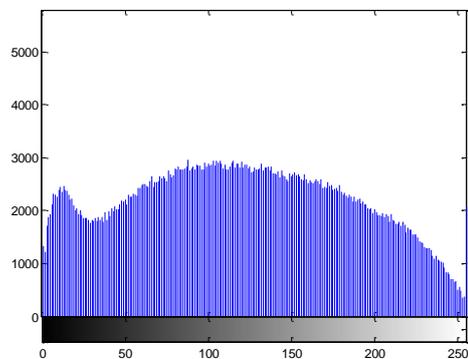
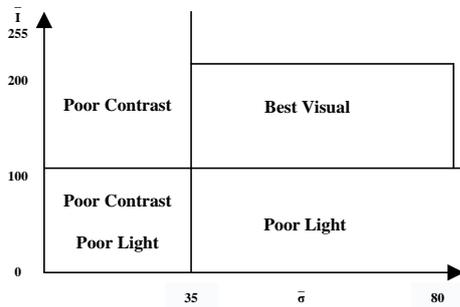Figure 5.   (a-d)Histogram of R channel from figure4(a) to (d)

Figure 6.   Distribution of visual effects

TABLE II.   STATISTICS OF $\bar{I} \times \bar{\sigma}$ (SUB-BLOCK SIZE 50×50)

	Original	Msrcr	He	This paper
Fig2(a-d)	622. 661 2	2.9e+03	3.8e+03	3.9e+03
Fig3(a-d)	764.5943	2.4e+03	2.9e+03	3.3e+04
Fig4(a-d)	876.1571	2.5e+03	3.3e+03	3.5e+03

We can see from Table 2 that quality evaluation result of the images which were processed by the algorithm in the paper was better than that of the traditional algorithm. $\bar{I} \times \bar{\sigma}$ value of the original image is less, and the corresponding brightness and contrast are worse. Quality evaluation data which were processed by the algorithm in the paper are improved generally compared with traditional algorithm, which indicates that the processing result of the algorithm in the paper were better than that of traditional algorithm.

Lastly, the operation time of the algorithm and objective image resolution area analyzed and the images are evaluated by using entropy function and Tenengrad function [13]. The results are shown in Table 3, Figure2, Figure 3 and Figure 4 are form top to bottom.

TABLE III.   SIMULATION DATA

	Original	msrcr	HE	Paper
entropy	7.5256	7.4830	7.6742	7.6844
Bre	4.3e+03	8.8e+03	9.6e+03	1.1e+04
Ten	1.2e+05	2.5e+05	2.7e+05	2.9e+05
Time		8.3156	134.3984	12.9643
entropy	6.9834	7.0565	7.1657	7.1455
Bre	2.2e+03	3.5e+03	2.7e+03	2.8e+03
Ten	5.6e+04	8.1e+04	6.9e+04	7.0e+04
Time		9.2847	129.3746	10.9874
entropy	6.9439	7.9056	7.3155	7.3103
Bre	2.2e+03	8.9e+03	6.1e+03	6.8e+03
Ten	5.7e+04	2.2e+05	1.5e+05	1.6e+05
Time		8.5769	117.9324	9.7193

From simulation images and Table 3, we can see that the defogging effect of HE algorithm and the algorithm of the paper is consistent basically, but the algorithm of the latter is 10% of HE algorithm, which indicates that the improved algorithm has certain practicability and effectiveness. Although the results of msrcr in Table 1 are good, the algorithm is not the restoration algorithm based on physical model, and the algorithm evaluated functions without color fidelity as the premise, which causes virtual height of the results.

IV.   CONCLUSIONS

The paper proposed a defogging algorithm for single image based on dark channel priority which can make rapid defogging on foggy images. For the disadvantage of HE algorithm, that is, much time is used to optimize transmission diagram, the paper applies a method combining adaptive median filtering and bilateral filtering to calculate dark channels, The dark channels are refined, and transmission diagram t estimated by the method is also refined without need to optimize, which reduces the complexity of algorithm significantly and overcomes the bottleneck of low speed for HE algorithm. The algorithm realizes rapid and high-quality defogging on single images in the condition of not influencing defogging effect. The improved algorithm has flexibility, practicability and effectiveness, and makes the reliability of outdoor visual system in foggy weather improve dramatically, so it has widely practical value.

REFERENCES

[1]   He K, Sun J, Tang X O, "Single image haze removal using dark channel prior", *Proc of IEEE CVPR 09.Washington: IEEE Computer Society*. pp. 1956-1963, 2009.
[2]   Ze Wang, Luxin Yan and Mingzhi Jin, "Optimal Stopping Condition for Iterative Deconvolution Method by The Measurement of Image Texture", *JDCTA,* Vol. 6, No. 15, pp. 109-119, 2012
[3]   S Shwartz, E Namer, Y Y Schechner. Blind haze separation. CVPR. *New York: IEEE Computer Society*, pp. 1984- 1991, 2011.
[4]   Wen-cheng Wang, Zhixiang Ji, Fengnian Guan, "Study on the Restoration Method from Motion Blurred Image", *AISS*, Vol. 4, No. 22, pp. 627 -632, 2012
[5]   S. G. Narasimhan, S. K. Nayar, "Contrast restoration of weather degraded images", pp.713-724, 2003.
[6]   Desai, N., Chatter jee, A., Mishra, S. et al, "A Fuzzy Logic Based Approach to De-Weather Fog-Degraded Images" , *In Proceeding of the Sixth International Conference on Computer Graphics, Imaging and Visualization*. pp. 383-387, 2009.
[7]   Ju Li, Xiao-ping Liu, Hui DU, "Research of Image Recognition based on Rough Set", *JDCTA*, Vol. 6, No. 9, pp. 141 -146, 2012
[8]   Fattal R. Single image dehazing //Proc of Siggraph'08. *New York: ACM*. pp. 1-9, 2008.
[9]   Deng Xiuqin, Xiong Yong, "Algorithm of Weighted Median Filtering Applied in Image Processing", *Computer Technology and Development,* No.3, 2009
[10]   Fang Shuai, Wang Feng, Zhan Jiqing, Cao Yang, Yuan Hongwu, Rao Ruizhong, "Synchronous Denoise and Recovery for Single Image", *Pattern Recognition and Artificial Intelligence,* pp. 91-100, 2012.
[11]   A. J. Preetham, P. Shirley, and B. Smits, "A practical analytic model for day light", *In SIGGRAPH,* pp.91-100, 2009.
[12]   S. G. Narasimhan, S. K. Nayar, "Vision and the atmosphere", *IJCV*, pp. 233-254, 2002.
[13]   Xu Guili, Liu Xiaoxia, Tian Yupeng, Cheng Yuehua, Li Peng, "A Evaluation Method of Image Resolution", *Infrared and Laser Engineering*, 2011.
[14]   K.N. Chaudhury, D. Sage, and M. Unser, "Fast O(1) bilateral filtering using trigonometric range kernels, " *IEEE Trans. Image Processing*, pp. 11, 2011.

[15] K.N. Chaudhury, "Acceleration of the shiftable O(1) algorithm for bilateral filtering and non-local means, " *arXiv*. pp. 1203, 2009.

[16] E. Hsu, T.Mertens, S. Paris, S. Avidan, and F. Durand. "Light mixture estimation for spatially varying white balance, " *In SIGGRAPH*, pp. 1–7, 2008.

[17] Shuai Fang, Jiqing Zhan, Yang Cao, Ruizhong Rao, "Improved Single Image Dehazing Using Segmentation", *In IEEE International Conference on Image Processing, Sep*. 2010, pp. 3589-3592.

[18] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski, "Deep photo: model-based photograph enhancement and viewing, " *International Conference on Computer Graphics and Interactive Techniques*, pp. 1–10, 2008.

[19] N. Hauti`ere, J.-P. Tarel, D. Aubert, and E. Dumont, "Blind contrast enhancement assessment by gradient ratioing at visible edgese," *Image Analysis & Stereology Journal*, 27(2) pp. 87–95, June 2008.

[20] R.T. Tan, "Visibility in bad weather from a single image, "*IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1–8, 2008.

[21] Wang, Z. and Simoncelli, E. P., "Translation insensitive image similarity in complex wavelet domain," *in Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing]*, Mar. 2005.

[22] Provenzi E, Gatta C, Fierro M, etal, "A spatially variant white-patch and gray-world method for color image enhancement driven by local contrast" , *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2008, 30(10) pp. 1757-1770.

Vulnerability Evaluation of Multimedia Subsystem Based on Complex Network

Xiaoling Tang

Institute of Higher Education Research, Jilin Business and Technology College, Changchun, China
E-mail:abc0863@21cn.com

Abstract—IP multimedia subsystems bring huge security threats to the IMS-based next-generation networks although they bring IMS-based network integration and business ability enhancement. Therefore, the research on the vulnerability of IP multimedia subsystems attracts widespread attention recently. In this article, the security threats of next-generation core networks are discussed first. And then we analyze the vulnerability characteristics of IP multimedia subsystems. In order to evaluate the vulnerability of IP multimedia subsystems effectively, this paper further applies the complex network theory to construct the network model of IP multimedia subsystem and propose the vulnerability assessment indicator by adopting entropy model. Finally, some simulation experiments and result analysis are provided to our evaluation model. Both theoretical analysis and numerical simulations indicate that the proposed vulnerability measurement is a simple and effective new method to evaluate the overall vulnerability of this kind of IP multimedia subsystems. In addition, the proposed method provides a new idea and scientific basis for the vulnerability assessment of IP multimedia subsystems.

Index Terms—IP Multimedia Subsystem, Network Safety, Vulnerability Assessment, Complex Network Theory, Entropy Model

I. INTRODUCTION

Many real-life networks such as the Internet, power grids and other self-organizing networks exhibit complex features [1, 2, 3]. These networks are important components of our daily lives. Their safeties can seriously impact the economic development and social stability. Therefore, the related researches on the network safety have attracted increasing attention in recent decades [4, 5, 6]. For computer networks, their security is more and more important. Therefore, it is urgent to develop a suite of measurable and quantitative metrics to evaluate their vulnerability and further propose schemes to protect them. Especially, IP multimedia subsystem is an important concept proposed by the most authoritative industry organizations 3GPP in the field of 3G mobile communications in March 2002 [7, 8]. As the core of the next generation networks, its function entity is numerous, and the vulnerability distribution is complicated [9, 10, 11]. Only when the vulnerability of IP multimedia subsystems can be evaluated effectively, we can find out the key factors which lead to the security performance threats of IP multimedia subsystems and control the

source of security threats within IP multimedia subsystems effectively [12, 13]. Furthermore, we can obtain the objective evaluation according to various vulnerability impacts on IP multimedia subsystems. And then the resilience of IP multimedia subsystems will be improved and enhanced.

Recently, 3GPP and 3GPP2 Organizations have proposed the security standards on IP multimedia subsystems. However, the study on the safety of IP multimedia subsystems is still in its infancy, which is insufficient to meet the security needs. At the same time, it is difficult to follow the network safety standard because the actual configuration of IP multimedia subsystems needs to consider the actual conditions and the specific network needs. Furthermore, in the operation and maintenance process of IP multimedia subsystems, there exist other factors, such as the human negligence and mistakes and so on. All these will lead to security threats of IP multimedia subsystems. In order to find out the weaknesses of IP multimedia subsystems, the source of safety threats and the possibility of leaked information, the vulnerability assessment of IP multimedia subsystems are very important and demanding to solve. In this way, the protection program of IP multimedia subsystems can be obtained. On this basis, we can further select and determine the safety control measures to avoid transfer or reduce the losses.

The early IP multimedia subsystems are the first step in the migration path. They tolerate the different known limitations of security which could be challenging to settle in real-life operations. However, the early IP multimedia subsystem is omit consciously but not without emphasizing the importance of considering the sharply decreased security [7]. Furthermore, the vulnerability of IP multimedia subsystem implementations and high level service provider security demands were examined. And then an implementation method to provide the desired degree of safety for IP multimedia subsystem deployments was defined [14]. Considering that communications service provider must determine how to handle the influences of security flaws, a new method was emphasized in reference [15], with examples to determine the influence of safety-related malfunctions and the key design elements which are related with modeling service reliability. However, the safety research results on IP multimedia subsystem are relatively lacking in spite of the core position of IP

multimedia subsystem in next-generation networks. So far, the vulnerability evaluation technology of domestic and international mainstream mainly consist of network survivability analysis techniques [16], attack tree technology [17], graph-based network vulnerability analysis techniques [18, 19]. The above technologies in solving the problem of vulnerability have some defects and shortcomings. Actually, the vulnerability assessment of IP multimedia subsystem should be a complex nonlinear problem since the vulnerability of IP multimedia subsystems is uncertain.

Considering that IP multimedia subsystems are in the core position in the next generation networks and the existing problems, in this article, the security threats of next-generation core networks are discussed and the vulnerability characteristics of IP multimedia subsystems are analyzed first. And then we use complex network theory to build a network model of IP multimedia subsystem. According to the characteristics of IP multimedia subsystem networks, a new solution for the vulnerability evaluation of IP multimedia subsystem networks is put forward on the basis of the entropy model. Finally, we provide some simulation experiments and result analysis to proof the effectiveness of our evaluation model. Both theoretical analysis and numerical simulations indicate that the proposed vulnerability measurement is a simple and effective new method to evaluate the overall vulnerability of this kind of IP multimedia subsystems. In addition, the proposed method presents a new solution method for the vulnerability measurement of IP multimedia subsystem networks and provides basis for the establishment of security policy of IP multimedia subsystems.

II. VULNERABILITY OF IP MULTIMEDIA SUBSYSTEM AND IP MULTIMEDIA SUBSYSTEM NETWORK

A. Vulnerability of IP Multimedia Subsystem

The vulnerability research of computer networks can be divided into several problem areas which include vulnerability definition, vulnerability identification, and vulnerability analysis and vulnerability assessment. In general, the main purpose of the vulnerability assessment is the right vulnerability scoring or rating, the qualitative and quantitative analysis of the system vulnerability. The system can be a service, a computer on the network or the computer network. In the field of network security, the vulnerability assessment and the intrusion detection of systems, firewalls, virus detection are four elements in network security.

The research for vulnerability assessment of IP multimedia subsystems is still in its infancy, there are no systematic research results up to now. Additionally, it is difficult to obtain the factors of vulnerability assessment. Therefore, according to the current situation of vulnerability assessment of IP multimedia subsystems, the analysis method of vulnerability dimension structure is introduced in this paper first.

The proposed evaluation method considers the security reasons of IP multimedia subsystems mainly from malicious users or attackers perspective. Defense

measures before the attack on the IP multimedia subsystem networks are investigated. This method has a more practical significance to service providers as well as the majority of users because it focuses on the difficulty and the impact that an attacker exploits the vulnerability of IP multimedia subsystem networks. An important consideration is provided to material utilization cost and time utilization cost which are not directly associated with the vulnerability of IP multimedia subsystem networks, but actually are closely related to the factors of their vulnerability utilization. On the basis of consideration, we analyze the vulnerability characteristics of IP multimedia subsystems and further present the assessment method. The proposed vulnerability assessment of IP multimedia subsystems is considered starting with the possibility of vulnerability utilization first. Secondly, the consequence of vulnerability utilization is an important factor which is considered by attackers. Synthetically considering above two aspects, we obtain the following evaluation factors of vulnerability. The detailed representation is as follows.
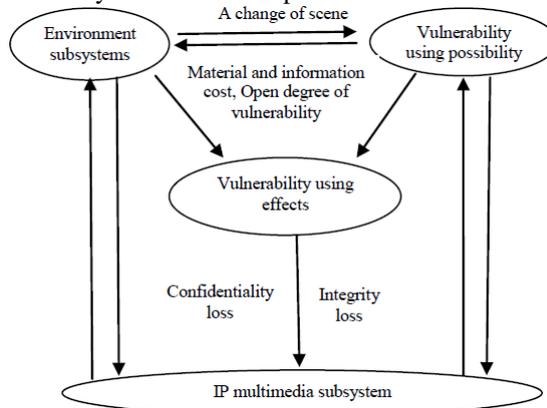

Figure 1. Index system of vulnerability assessment model.

IP multimedia subsystems are vulnerable to attacks and malfunctions. The most important reason is that there are many weaknesses in themselves. Once an attacker exploits these vulnerabilities, they are easy to succeed. The vulnerabilities of IP multimedia subsystems include the system vulnerability which exists initially and the potential vulnerability which is brought by the later increased safety measures. In other word, the vulnerability of IP multimedia subsystems refers to the number of loopholes in the system and the resulting damage degree to the potential threats. It is an important indicator to measure safety performance of an IP multimedia subsystem. In general, if the vulnerability degree of an IP multimedia subsystem is small, it is more secure. Here, the vulnerability of IP multimedia subsystems is divided into four categories: logic errors, system flaws, social engineering and strategy negligence. Each category in the above four categories can be subdivided into several subcategories. In order to be convenient to calculate them, here the factor set of vulnerability assessment and its corresponding subset of factors are designed. According to the actual situation, the weight is given according to various factors, as shown in Table 1.

TABLE I. OUR PROPOSED VULNERABILITY EVALUATION FACTORS OF
IP MULTIMEDIA SUBSYSTEM

Level indicators	Weights	Secondary indicators	The corresponding level of the weight of the higher level
Logic errors	0.25	System bug; Hardware bug	0.65; 0.35
System flaws	0.25	Firewall filtering rules; Security control measures	0.35; 0.65
Social engineering	0.30	Administrator service quality; personnel security awareness of IP multimedia subsystem	0.40; 0.60
Strategy negligence	0.20	Security policy defects; Implementation effectiveness of security policy	0.50; 0.50

B. IP Multimedia Subsystem Network based on Complex Network Theory

With the rapid development of IP multimedia subsystem, network consistency rule becomes a very active topic in network dynamics. Since small world networks and scale-free networks are proposed, complex networks have become the popular research topic in various disciplines. Some basic concepts and characteristic quantities such as the shortest path, degree distribution, clustering coefficient, betweenness and so on are put forward gradually to describe the topology characteristics and dynamic properties of complex networks. It is found that there are two basic characteristics which are large average clustering coefficient and shortest path. Similar with social relation feature in social network, the actual networks also include a community structure. That is, the inner connection between nodes of each community is very close. These characteristics have paved the way for the study of IP multimedia subsystems.

Based on the characteristic analysis of the vulnerability of IP multimedia subsystems, the architecture-based vulnerability approach is adopted to investigate IP multimedia subsystem networks by combining complex network theory. The architecture of is given by its probabilistic control flow graph. In this article, a graph G is defined by a non-empty set of nodes H and a nonempty set of edges L : $G = (H, L)$, where $H = \{v_1, v_2, \cdots, v_N\}$ is the set of nodes and $L = \{e_1, e_2, \cdots, e_M\}$ is the set of edges. The total number of nodes in the set H is N, and the total number of edges in the set L is M .

In an IP multimedia subsystem, information exchange behavior always occurs at the point. The main function of the subsystems is to connect two relatively independent components together. We will also regard any pair of points allow access as a link, as is shown in Figure 2.
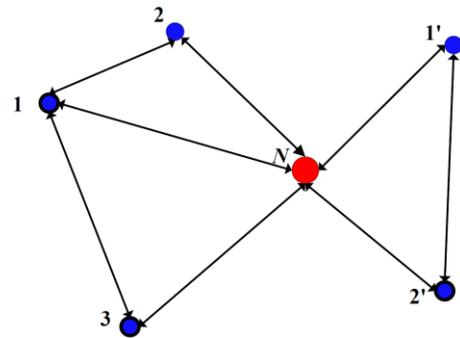

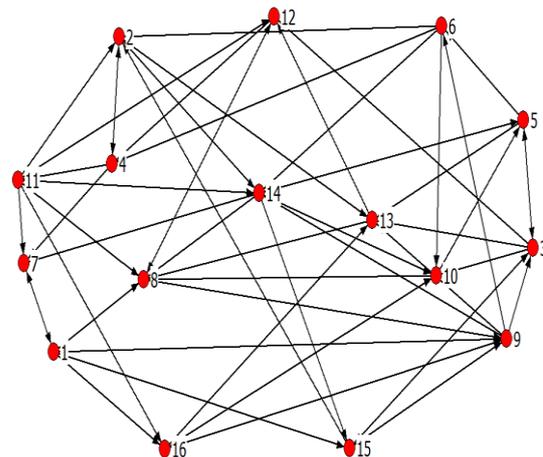Figure 2.   Information exchange link graph within IP multimedia subsystem network

Figure 3.   Virtual representation of IP multimedia subsystem network

In the following, we make ensure the components of IP multimedia subsystem network. Then the exchange relations of information and energy among components are investigated the between components. When there is the exchange of matter or energy between the two component nodes existing within the IP multimedia subsystem, an edge between them is determined. The Nodes represent the components of the application. The edges represent the flow of control among the components. Networks can be directed or undirected. Undirected networks are composed by edges connect a pair of nodes in a transitive fashion. Nevertheless, directed networks are composed only by directed edges connecting a pair of nodes through a given direction. In general, directed edges are showed by an arrow representing the direction of the connection between a pair of nodes. Here, we will extract IP multimedia subsystem into a directed network. In order to obtain the corresponding adjacency matrices of IP multimedia subsystem, we collect data and investigate the relations among component nodes and then obtain the adjacency matrix. Figure 3 represents the probabilistic control flow graph of an IP multimedia subsystem network by using UCINET 6.0, which consists of 16 nodes.

III.   PROPOSED SCHEME

For IP multimedia subsystem network, the two main indicator of the information flow is the timeliness and accuracy of the information transmission, as well as

information security, namely the network vulnerability. Considering these three aspects, we use the entropy to construct the vulnerability assessment model of IP multimedia subsystem network.

Based on the above analysis of the vulnerability of IP multimedia subsystem networks, the uncertainty of microscopic state of IP multimedia subsystem networks is described by the entropy in the following. It is related with the number of the possible microscopic states under macroscopic states and also with the occurrence probability of the different microscopic states. The entropy reflects the reality connection between the nodes within the network. The maximum entropy E_m represents the maximum value of the network entropy, which represents the most chaotic state of the IP multimedia subsystem network. The ratio of the present network entropy and the maximum network entropy reflects the unorganized extent of the IP multimedia subsystem network. Therefore, we define the ordered degree of an IP multimedia subsystem network as follows.

$$R = 1 - \frac{E}{E_m} \quad (1)$$

The value of R characterizes the degree of organization of the network. It is larger, which means that the ordered degree of network is higher and their organization is more efficient.

In the following, we define the ordered degree of a network based on the accuracy and network survivability during information transmission of IP multimedia subsystem network respectively in this article. And then we synthetically consider the two aspects. In this way, network evaluation model is proposed on the basis of the consideration for the accuracy of the response and the network anti-destroying ability.

Network quality reflects the accuracy of the information flow between the nodes in the network. The quality entropy characterizes the size of the uncertainties of information quality. The total quality entropy is defined as follows.

$$E_1 = \sum_{k=1}^{N} E_1(k) \quad (2)$$

$$= -\sum_{k=1}^{N} p_1(k) \log p_1(k) \quad (3)$$

where $p_1(k)$ is node k to achieve the quality of the micro-state probability? The expression of $p_1(k)$ is as follows.

$$p_1(k) = \frac{d_k}{\sum_{k=1}^{N} d_k} \quad (4)$$

where d_k is the number of nodes which is directly linked to node k. It is also the above-mentioned node degree in IP multimedia subsystem network.

The maximum quality entropy of IP multimedia subsystem network is defined as follows.

$$E_{1m} = \log \sum_{k=1}^{N} d_k \quad (5)$$

Thus, the quality ordered degree is as following.

$$R_1 = 1 - \frac{E_1}{E_{1m}} \quad (6)$$

Cluster coefficient C represents the cluster level among nodes in the network. It describes the aggregation state of nodes within network. For the cluster coefficient of any node k, the following definition is adopted:

$$C_k = \frac{2E_k}{d_k(d_k - 1)} \quad (7)$$

where E_k is the number of actual existing edges among neighboring nodes of node k. $\frac{d_k(d_k - 1)}{2}$ is the maximum connection number among all neighboring nodes of node k.

Thus, the average cluster coefficient of the entire IP multimedia subsystem network is as follows.

$$C = \frac{1}{N} \sum_{k=1}^{N} C_k \quad (8)$$

Obviously, $0 \leq C \leq 1$.

Anti-destroying ability reflects the stability size of the network structure under different attack strategy. Anti-destroying entropy describes the uncertainty size of the network structure anti-destroying ability. Anti-destroying entropy of node k characterizes the uncertainty size of the anti-destroying ability of node k, which is defined as following.

$$E_2(k) = -p_2(k) \log p_2(k), \quad k = 1, 2, ..., N \quad (7)$$

where $p_2(k)$ is the microstates anti-destroying probability of node k and the following expression of $p_2(k)$ is adopted:

$$p_2(k) = \frac{C_k}{\sum_{k=1}^{N} C_k} \quad (8)$$

According to Eq. (7) and Eq. (8), we define that the anti-destroying entropy of IP multimedia subsystem network is as following.

$$E_2 = \sum_{k=1}^{N} E_2(k) \\ = -\sum_{k} p_2(k) \log p_2(k) \quad (9)$$

The maximum anti-destroying entropy of IP multimedia subsystem network is defined as follows.

$$E_{2m} = \log \sum_{k=1}^{N} C_k \qquad (10)$$

Thus, the quality ordered degree is as following.

$$R_2 = 1 - \frac{E_2}{E_{2m}} \qquad (11)$$

The vulnerability measurement is denoted by R. In order to consider the two main factors of the accuracy of the information transmission and information security synthetically, the following vulnerability measurement of IP multimedia subsystem network is adopted based on Eq. (6) and Eq. (11).

$$R = \mu R_1 + \lambda R_2 \qquad (12)$$

where μ and λ are the weight coefficients respectively.

In E_m is the same case, if R is larger, the ordered degree of IP multimedia subsystem network is higher. In other words, IP multimedia subsystem network is more efficient. Therefore, the value of R could characterize the vulnerability of IP multimedia subsystem network well.

To make a meaningful comparison, we introduce another definition to discuss the vulnerability of the IP multimedia subsystem networks. We investigate their global connectivity. A parameter G is proposed to measure the global connectivity, which is defined as follows.

$$G = 1 - \frac{N'}{N} \qquad (13)$$

where N is the number of the nodes which can maintain the normal operations before cascading breakdowns, and N' is the number of the nodes of the whole network which can maintain the normal operations at present.

We define that the load of node k is as following.

$$L_k = C_B(k) = \frac{\sum_{i \neq j \in H} \dfrac{\delta_{ij}(k)}{\delta_{ij}}}{N(N-1)} \qquad (14)$$

where δ_{ij} is the number of all the shortest paths between node i and node j, $\delta_{ij}(k)$ is the number of the shortest paths through node k between node i and node j. The capacity of a node represents the maximum load that the node can tolerate. In IP multimedia subsystem networks, the capacity of a node is limited by the corresponding cost. Here, we assume that the capacity C_k of node k to be proportional to its initial load, i.e.,

$$C_k = (1+\alpha)L_k(0)$$
$$= (1+\alpha)\frac{\sum_{i \neq j \in H} \dfrac{\delta_{ij}^0(k)}{\delta_{ij}^0}}{N(N-1)} \qquad (15)$$

where the parameter α is the tolerance parameter which reflects the resistance to the attacks and malfunctions, According to the actual situation, $L_k(0)$ is calculated by Eq.(14) and the capacity C_k of node k is calculated by Eq.(15) in our simulations.

IV.   NUMERICAL RESULTS

In the simulations, we choose $\mu = 0.5$ and $\lambda = 0.5$. We use Matlab2010 as a programming language. In order to implement our model, the Floyd Warshall algorithm is utilized in this letter. This algorithm follows a dynamic programming method to calculate the shortest paths between all vertices of an IP multimedia subsystem network. Based on the proposed vulnerability assessment method of IP multimedia subsystem network, the IP multimedia subsystem network model including 16 nodes will be applied to illustrate the vulnerability of the IP multimedia subsystem networks which is subjected to malicious attacks. The betweenness for a node is the number of shortest paths passing through the node. We adopt the highest betweenness node-based attack to explore the vulnerability phenomenon of the IP multimedia subsystem network. If there are some nodes have equivalent betweenness, we choose the node with the highest degree as the attacked node.
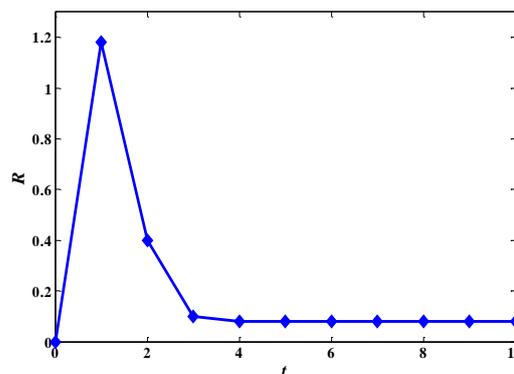

Figure 4.   The relation between the proposed measurement R and the runtime t by removing the node with highest betweenness under our model in the case of $\alpha = 0.0$.
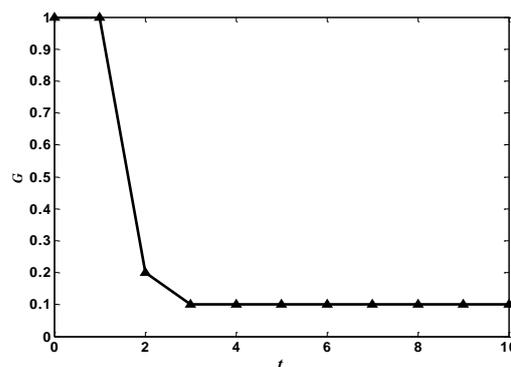

Figure 5.   The relation between the damage size G and the runtime t by removing the node with highest betweenness under our model in the case of $\alpha = 0.0$.
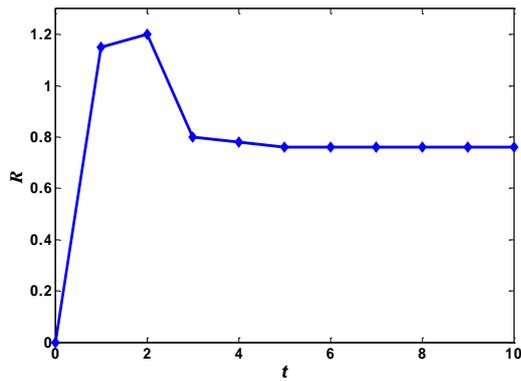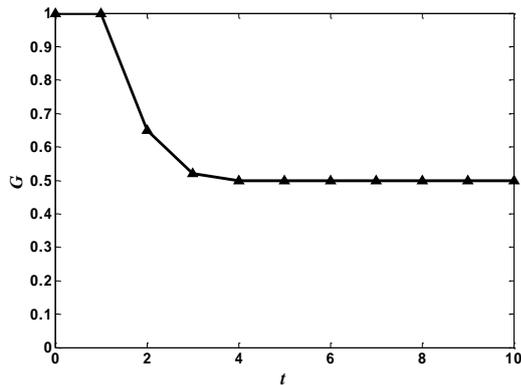
Figure 6.   The relation between the proposed measurement R and the runtime t by removing the node with highest betweenness under our model in the case of $\alpha = 0.1$.

Figure 7.   The relation between the damage size G and the runtime t by removing the node with highest betweenness under our model in the case of $\alpha = 0.1$.
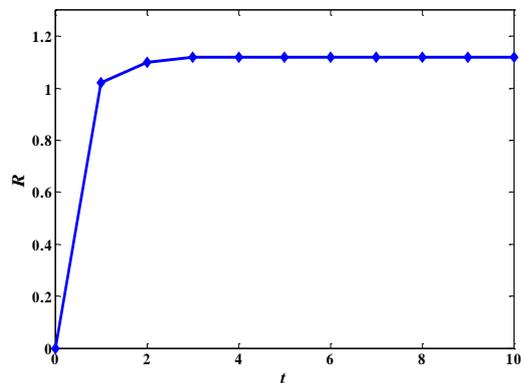

Figure 8.   The relation between the proposed measurement R and the runtime t by removing the node with highest betweenness under our model in the case of $\alpha = 0.3$.

In Figs. 4-9, the dynamical behaviors of our proposed vulnerability measurement based on entropy and the damage size during failure propagation are reported for an IP multimedia subsystem network of 16 nodes at the cases of different values of the parameter α. To make a meaningful comparison, From Figs. 4, 6 and 8, we can see that the value of R for a larger cascade increases much more sharply than that for a smaller one even at the

very early stage of malfunction propagation when the node with the largest betweenness is attacked. Therefore, the proposed vulnerability measurement of IP multimedia subsystem network can help us to identify the larger damage at the very early stage of malfunction propagation. It is of great significance in the prevention of major accidents in IP multimedia subsystems. It is because that the major accidents are not distinct from the smaller malfunctions on the entire scales at the early period of accident propagation. In addition, as shown in Figs. 4, 6 and 8, we also find that the value of R increases to a peak value and then decreases to a stationary value during the process of malfunction propagation. In other words, it increases to a maximal value and then maintains almost unchanged. This result shows that our method can make us find the malfunctions and dissolve locally in time. Thus, the malfunctions from propagating throughout the IP multimedia subsystem network are prevented effectively.
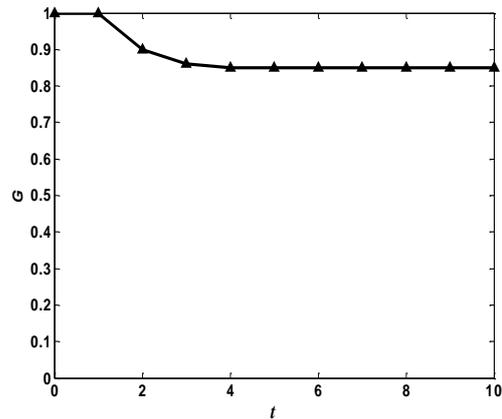

Figure 9.   The relation between the damage size G and the runtime t by removing the node with highest betweenness under our model in the case of $\alpha = 0.3$.

Figs. 5, 7 and 9 depict the dynamical evolving behaviors of damage size during the malfunction propagation in the IP multimedia subsystem network of 16 nodes, at the cases of different values of the parameter α. In many IP multimedia subsystem networks, the parameter α is severely limited by the system cost. Therefore, our simulations adopt different values of parameter α to investigate the IP multimedia subsystem network. We examine the impact of different node capacity of α on the IP multimedia subsystem networks, as is shown in Figs. 4-9. The comparison results of the dynamics of the value of R and damage scale R during malfunction propagation are shown in Figs. 4-10. We can see that the value of parameter α has significant influence on the malfunction propagation in IP multimedia subsystem network. Since the value of parameter α is limited by the cost of an IP multimedia subsystem, we should construct the load distribution of the IP multimedia subsystem networks as homogenous as possible to decrease the vulnerability against the major incidents. By simulations, we see that our results are in agreement with easily the real-life IP multimedia

subsystem. For example, in an IP multimedia subsystem, the more equal the exchange information capacity is, the more stable the IP multimedia subsystem becomes. Consequently, our proposed vulnerability evaluation based on entropy can be used as an indicator to measure and control the vulnerability of IP multimedia subsystem networks.

V.   CONCLUSION

IP multimedia subsystem networks provide a powerful IP session-based core for advanced services in next-generation networks. They must be suitable to meet customer needs and service provider. IP multimedia subsystems are designed to be a generic service enabler platform which includes the availability requirements of the present telecommunication operators' networks. In order for facilitate the operations of this complex and costly architecture, we will a plenty of attractive services obtaining high user acceptance. The vulnerability of the IP multimedia subsystem user equipment is also a serious safety threat. At present, the vulnerability study of IP multimedia subsystems in its infancy, no more influence results are put forward by authoritative experts or institutions. Therefore, it is not easy for us to quantify each evaluation attributes of the vulnerability for IP multimedia subsystem networks. There is too little safety and too many restrictions imposed upon the users and customers that make operating adequate security of IP multimedia subsystems such a grand challenge.

The entropy is initially introduced as a thermodynamic concept. It is a measure for the system disorder. The entropy is widely used due to its unique content and penetration. Recently, as a physical description of complex system structure, the entropy is more and more attention and becomes an important research tool for complex systems in the theory of complex systems [20, 21]. The macro significance of entropy is a measurement for the uniformity of the system energy distribution. The change direction of the complex system state can be expressed by the entropy. The energy distribution is more uniform, the greater the entropy is. Based on this, in order to avoid subjective weighting method of uncertainty caused by the vague and controversial and determine the scientific and reasonable vulnerability evaluation method, we introduce the concept of entropy to explore the vulnerability of IP multimedia subsystem networks.

In this paper, on the basis of the emergence complex network theory in recent years, we analyze complex features of IP multimedia subsystem networks and explore the factors of the vulnerability utilization for IP multimedia subsystem networks. Based on the vulnerability analysis, we introduce the complex network to build a network model according the features of IP multimedia subsystems. The concept of the network entropy is proposed to quantify the vulnerability of IP multimedia subsystem networks. Finally, we provide some simulation experiments and result analysis to confirm the effectiveness and the practicality of our evaluation model. Both numerical simulation results show that the proposed vulnerability measurement is a simple and effective new method to evaluate the overall vulnerability of IP multimedia subsystems. The proposed method provides a new solution method for the vulnerability measurement of IP multimedia subsystem networks and also provides basis for the establishment of security policy of IP multimedia subsystems.

REFERENCES

[1] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, pp. 1025, 2010.

[2] P. S. Koutsourelakis, "Assessing structural vulnerability against earthquakes using multi-dimensional fragility surfaces: a Bayesian framework," *Probab. Eng. Mech*, vol. 25, pp. 49-60, 2010.

[3] M. E. J. Newman, "The structure and function of complex networks," *Siam Review*, vol. 45, pp. 167-256, 2003.

[4] A. A. Moreira, J. S. Andrade, H. J. Herrmann, and J. O. Indekeu, "How to make a fragile network robust and vice versa," *Physical Review Letters*, vol. 102, pp. 018701, 2009.

[5] Y. Q. Jiao, J. Weir and W. Q. Yan, "Flame detection in surveillance," *Journal of Multimedia*, vol. 6(1), pp. 22-32, 2011.

[6] J. L. Wan and Y. H. Liu, "Multi-regions texture substitution for image and video," *Journal of Multimedia*, vol. 7(6), pp. 394-400, 2012.

[7] Third Generation Partnership Project (3GPP). Security aspects of early IP Multimedia Subsystem (IMS) (Release 8).*TR 33.178 V1.0.0, 2008.*

[8] J. L. Wan and Y. H. Liu, "Signalling cases and QoS management within TISPAN NGN residential environments," *Journal of Multimedia*, vol. 2(2), pp. 28-36, 2007.

[9] A. Barnawi, A. H Altalhi, M. R. J. Qureshi and A. I. Khan, "Evaluation of IP multimedia subsystem based mobile mass examination system," *I. J. Computer Network and Information Security*, vol. 4, pp. 1-12, 2012.

[10] J. H. Lee, Y. B. Huh, S. W. Woo and J. S. Sohn, "Routing mechanism for VoIP emergency calls in IP Multimedia System" *Advanced Communication Technology* (*ICACT*), 2010 *The 12th International Conference on*. Vol. 1, pp. 392-395, IEEE, 2010.

[11] C. Y. Chen, K. D. Chang, J. L. Chen and H. C. Chao, "Extending emergency services coverage in cooperative IMS networks," *International Journal of Autonomous and Adaptive Communications Systems*, vol. 5(4), pp. 417-434, 2012.

[12] P. Bellavista, G. Cardone, A. Corradi and L. Foschini, "The Future Internet convergence of IMS and ubiquitous smart environments: An IMS-based solution for energy efficiency," *Journal of Network and Computer Applications*, vol. 35(4), pp. 1203-1209, 2012.

[13] A. Berger, I. Gojmerac and O. Jung, "Internet security meets the IP multimedia subsystem: an overview," *Security and Communication Networks*, vol. 3(2-3), pp. 185-206, 2010.

[14] E. E. Anderlind, D. W. Faucher, E. H. Grosse, et al., "IMS security," *Bell Labs Technical Journal*, vol. 11(1), pp. 37-58, 2006.

[15] H. Pant, A. R. McGee, U. Chandrashekhar and S. H. Richman, "Optimal availability and security for IMS-based VoIP networks," *Bell Labs Technical Journal*, vol. 11(3), pp. 211-223, 2006.

[16] D. Farmer and W. Venema, "Vulnerability assessment franwork," *NewYork: CIAO Publication*, pp. 125-129, 2006.

[17] C. R. Ramakrishnan and R. Sekar, "Survivability analysis of network specifications," *New York: CIAO Publication*, pp. 155-159, 2005.

[18] L. Wang, S. Noel and S. Jajodia, "Minimum-cost network hardening using attack graphs," *Computer Communications*, vol. 29(18), pp. 3812-3824, 2006.

[19] S. Jajodia and S. Noel, "Topological vulnerability analysis," *In Cyber Situational Awareness*, Springer US, pp. 139-154, 2010.

[20] M. Caserta and M. C. Nodar, "A cross entropy based algorithm for reliability problems," *Journal of Heuristics*, vol. 15(5), pp. 479-501, 2009.

[21] F. Altiparmak and B. Dengiz, "A cross entropy approach to design of reliable networks," *European Journal of Operational Research*, vol. 199(2), pp. 542-552, 2009.

Global Non-linear Dimensionality Reduction Algorithm in Image Sets

Xinbo LAN, Dongrui LI, Shanyou YANG
Department of Computer, Guangdong AIB Polytechnic College, Guangzhou, China
Email: sinboland@163.com

Qian CHEN
College of Science and Engineering, City University of Hong Kong, Hong kong, China

Abstract—**Most of the real world image is in the area of all possible pictures level instance of low dimensional characteristics. Know this low dimension to understand many image set is a key first step, has been applied to the reduced-order measures in different jurisdictions, including facial recognition, machine learning, image retrieval and handwritten character recognition. Traditional evaluation of statistical relationship distance global dimension reduction arithmetic. This assignment optimization assessment measures the relationship between the proposed a global data--GODR. The relationship between the measurement in the implementation of the data set and they are the intersection point. As the point of view, global dimension reduces the built-in implementations. Compared to worldwide arithmetic, should the divisor, and have more useful result. Facial expression implemented the change the efficiency of the arithmetic is verified by the experiments.**

Index Terms—**Image, Dimensionality Reduction, Face Recognition, Image Retrieval**

I. INTRODUCTION

Behaviors in the video assume some forms: people walking, cycling and jogging, bird, fishing, swimming, kangaroos jumping, water fountain spray, heartbeat, contrast agent by the tissue, MRI marker lines slowly disappear. The actions constitute the semantic component of the videos. Achieve or withdrawal represent these behaviors is the important step to understanding the semantics of videos. This also applies to boar den image data sets, the order is not instantaneous. Specific and particular object image may range according to posture, lighting, all kinds of noise, and misunderstandings in pictures obtain processes. Understand or extract indicates this is the most important thing to effectively indicating image data.

What action are the appropriated presentations of the video or image collection changes? Model in this article of natural happening, Vase Thompson thinks carefully study the deformation is crucial:

"Among the very big part of our important task is to compare the form and in their precise definition, deformation complex plans may be a easy to understand the phenomenon, although this figure itself is impossible to leave and undefined analysis."

There are many possible targets in a series of contrast images, i.e. the cluster - a partitioned data set similar set of discontinuous elements and deterioration - discloses a continued mode data set may have originated. As for the image, it could be regarded a high-dimensional data point (each pixel position corresponds to some data set components), and regression skills are often not effective, we have explained the "curse of dimensionality" [1]). But, many different images and video can just be because a tiny number of aspects of freedom or setting close to some of these sub-sitting or coherent low-dimensional manifolds in high-dimensional space.

The principal objectives of manifold researching aims to restore the lower dimensioned manifold framework observed for the higher dimensioned data, and gain the objective of dimension reduce process. This is the most significant machine learning methods, with far-reaching applications and research in an area of computer utilizations [2-5]. A manifold studying algorithm is on the base of angle measurement and weight, so the map for the higher dimensioned data per hyper sheered and maintaining internal communities through local conformal mappings. For a local tangent space alignment (LTSA) arithmetic [3], and PCA can be used as the local tangent space [5, 10], so the small coordinates of the projection to a lower dimensioned space the global affined transformation to obtain local coordinates entire low-dimensional space coordinates. But, using data sets LTSA should be smooth, different. These are the ideal suggestions, higher request data. Really, the tangent space alignments are sometimes suitable geometric form data. Higher dimensioned data are always complained, the devotion of the tangent space can be enhanced when the datasets away from the tangent space. To avoid this problem, in [4] ISOMAP geodesic distance arithmetic to use, it requires a lot of computation in dealing with large samples.

This article pays attentions to dimension reduced process arithmetic aiming to know differences in behavior or image statistics sets. The main contribution of this paper is as follows:

We present article worldwide perfect dimension reduced process arithmetic. This update arithmetic to measure the relationship between the sample and the

orthogonal projection data set, achieving global embedding dimension reduction based on the most optimal angle.

GODR is a worldwide threads measures, and there is a good degree of public and PCA. It applies the angle between the sample is measured; it is not the main component of sensitive data.

GODR carry eigen decomposition in concentrated sample, to avoid overflowing the characteristics calculated using principal component analysis is very high dimensional models.

Facial description acknowledges experiments confirmed, GODR have better recognition consequences than the PCA.

II. DIMENSION REDUCTION METHODS

In the beginning, we offer a range of natural picture sets, there is a manifold framework. These models will highlight some of the measures, natural images are often different, and provides a graphic insight some usage dimension reduction picture. For everyone, we express topological manifold dimensions and generate an image.

Another proper case of point, the change is expressed in Fig. 1. In this data set is a bird in the sky image), there have 2 aspects of freedoms: to incite rigid information as a bird wings, and a rigid deformation of birds past by cameras. The second aspect of freedom appear in the right column of this figure, pun bird flying in the same part of the cycle, but that the opposite from different directions.

This image topological manifold is specific. This is a photograph that is the most natural parameters of the cycle parameters codification which part of the image capture flight period, or a thread parameter encoded far away for the camera.

Figure 1.   Samples of a picture manifold produced through deformation due to the flap.

Produced the number and nature image sets lower dimensional, however the non-linear structure as one shown below, this is very natural to regard arithmetic's to automatically discover the framework. In the past decade, a surprising amount of the arithmetic has been found, recently concluded that most of the highlighted widespread focusing on the region. In the part, we will provide most of other outstanding this arithmetic and demonstrate what options are important for this arithmetic is efficient pictures data.

Dimension reduction, and not threads dimension reduce process, is a corresponding principal component analysis, which seeks to get a lower dimensioned data sets parameterized nonlinear manifold lying in the higher dimensioned area. Generally, the data could be regarded as a point out of a roll. The coordinates of these points can be three (or more), there is a potential for two-dimensional representation. Dimensionality reduction goal is to automate to inductive the models.

When an initiate statistics is located in an area, and dimension reduction aims to discover coordinates for each part in a tiny dimensioned field d.

A given input setting, which is the set of them, for more dimensions D, a parametric study generates a mapping f: retain some of the attribute structure X.

Not thread dimensionality reduce process of this issues for decades the most widespread interest. Nevertheless, for this investigation, we will begin to discuss the new focus on the topic when triggered by two arithmetic's (iso metric feature mapping (Isomap) and Locally Linear Embedding (liquid - liquid extraction method) are also given. Almost the identification time, the most famous Desktop computer store or processed power required from the analysis of big pictures and vided opens a fertile field of application of these methods. Liquid - liquid extraction are two ways Isomap, reflect the different measures to define more specifically the above, we these arithmetics are discussed first, and then give an overlook of the option.

A. Isomap:

Is a dimension reduction arithmetic Isomap retained the geometric characteristics of the input statistics. In fact, the objective is to regain an Isomap isometric,

$$f : x \rightarrow y \tag{1}$$

for $x \rightarrow R^D, y \subset R^d$, and $d \ll D$ where, from all pairs of these samples $X_i, X_j \in X$,

$$| X_i - X_j |_{shortest\ path} = | Y_i - Y_j |_2 \tag{2}$$

I.e., Isomap regains an embedded can be approximately same to the distance between the smallest road distances (defined in the figure neighbors in an initiate space). Here, we express the most significant steps of the arithmetic.

The last step you need to embed MDS Isomap use. MDS can be the symbol of a matrix eigenvalue issue is transformed into absolute co ordinate's pair wise distance in some (usually low) dimensional space.

There is original Isomap arithmetic and liquid - liquid extraction of data sets, such as the work of "cake" in Figure 2. but, the data with higher curve, intersection and approaching self relation and convex sampling (in manifold space) lead two Isomap and liquid - liquid extraction failed. Table I show most of the recent changes, help to solve related backwards.

ST-Isomap is utilized to data in a time components, i.e., work of videos, or by modifying the structure of the local community and the distant matrix to narrow the distance

between two adjacent space and time points. We are the landmark deal to leave Isomap right speed accuracy using only the set of the sample built-in steps. There has given the certain constraints in a sample of data points.

TABLE I. DIMENSION REDUCED ARITHMETICS

arithmetic Class	Samples
Isomap variants	ST-Isomap
	Continuous Isomap
	Landmark Isomap [11]
	Conformed Isomap [11]
Charting	Manifold picturing [12]
	Nonlinear CCA and PCA [13]
Graph Spectral Methods	Laplacian Eigenmaps [14,15]
	K-Eigenmaps [16]
	H-Eigenmaps [17]
	Locality Maintaining Projections [18]
Supervised Methods	Native Fisher Built-in[19]
	Supervised LLE[20]
Other things	Diffused pictures[21]
	Manifold Tangent Studying[22]
	Proximity pictures[23]
	Semidefinite Built-in[24]
	Stochastic Neighboring Embedding [25]
	Native Smoothing[26]

Another arithmetic, semi-definite Embedding (SDE), and Isomap goal is to offer an isometric built-in. actually, the final step, using the same embedded. MDS main difference can be within the establishment of not difference matrix among all the pair of input samples. SDE is the application of semi-definite learning the kernel matrix. This method does succeed for non-convex, can be correctly expressed as Isomap with the parameter "P-shaped."

Self-organizing map, also called Kohonen attribute map, available at many of the arithmetic's in Tab. I indicated as the visualized method from higher dimensioned statistics, using a self-organizing map dimension reduction is later found the. Some go along the popular workspace of artificial intelligence net training using competitive learning to find a lower dimensioned built-in of input statistics.

Mapping method represents a higher dimensioned framework as a group in the grinding "Charts" or "pictures." Chart size can be fixed or can be extended to a number of assumptions are violated, i.e., native planarity. Liquid - liquid extraction compared, Isomap, many of the measures cannot offer a worldwide consistent set of input data parameterization, but the local area parameterization.

All the arithmetic expressed is essentially unsupervised; nevertheless, there is a relevant degree of full overviewed dimension reduce arithmetic becomes fluid - liquid extraction. The arithmetic often generally assumes that per sample is expressed as a native retention integration neighboring, however include additional constraints measures with the bias final solution. The following section expresses methods for their professional these usual arithmetic is a valid image.

III. GLOBAL OPTIMIZED DIMENSIONALITY REDUCTION (GODR)

Described in the previous section, arithmetics have been developed for the general problem of dimensionality reduction. These can be directly applied to the image data set, the picture is considered as samples in a higher dimensioned picture area, per coordinate representative a pixel intensity value.

Since a group of image defines a set of projection in this very higher dimensioned area, dimension reduction can be applied directly to the point set; early literature provides results that a simple set of images. Three key challenges limit the effectiveness of direct application of dimensionality reduction arithmetic for image data in this way. First, the natural image is often a curved manifold in this space. Second, many dimensionality reduction arithmetic's do not give natural tool for projects and project a new image to the reinsurance manifold. That is, once the nonlinear dimensionality reduction is applied to a group of pictures, no natural tool to find the coordinates of embedding a new image, there is no natural tool with arguments on the manifold and generates a sample image. Third, the natural image data sets often arise from manifolds with cyclic parameter, which is a challenge many dimensionality reduction arithmetic.

Isomer estimation arithmetic is based on the distance between the images. Change this distance metric is a natural way to modify the behavior of this dimensionality reduction arithmetic. The most common distance metric is the Euclidean distance is reduced dimension. If the image is considered as samples in a higher dimensioned area, each dimension corresponding to the identity of a point, then the Euclidean distance corresponding to the calculated square root of sum of squares of differences in pixel intensity. However, the image usually changes because of movement and deformation of the objective on view. In these cases, other methods may be more faithfully from a different one pair of image capture.

The second major limitation is that most of the dimensionality reduction method defines a map for an initiate data set, where d is the low-dimensional embedding space. It is, the result is a map

$$f : x \rightarrow R^d \tag{3}$$

or not, as may be most convenient,

$$f : R^D \rightarrow R^d \tag{4}$$

This indicates that embedded dataset X is calculated, because the values are not explicitly defined. In addition, we are also problematic inverse map. As a projection, if Y can be not set at point definition, then either not well confined.

Although some general method proposed to calculate these "sample" forecast, which is that both in theory and practice, a huge challenge for nonlinear dimension reduced process skills. Generally speaking, these methods evaluate a new point embedded within the community by finding the master data set and interpolated embedded connection. Conversely, if we have a new embedded image coordinates, the coordinates for those generated by linear interpolation image is embedded in the vicinity.

In some problem areas, we are with additional information on the type of difference images. The case where the image is changed, because the Buddha

deformation arithmetic agreement made clear deformation in each manifold coordinates using free-form deformation. Recently arithmetic, seeking to change the shape of the brain characteristics as they age, explicit consideration of infinite-dimensional manifolds diffeomorphic to realize the potential is converted to a regression model describes a degree of freedom while ignoring others.

Global dimension reduction of traditional algorithm measurement distance relationship. This paper proposes a measuring method of global optimization, data GODR. The relationship is between the measurement data of concentration of samples and their orthogonal projection. As the Angle of optimization, achieving global dimension reduction, embedded. Comparing with the global methods, the method is more efficient effects. Face expression recognition efficiency of the algorithm is verified by the experiments.

In order to decrease errors that regular data brought aims to enable the distance minimum between the subspaces: $space(Q)$ and $space(\hat{Q})$.

Notes: the far way between 2 areas $space(Q)$ and $space(\hat{Q})$ can be $dist(Q,\hat{Q}) = \sqrt{1-\sigma_1^2}$, where $\cos\theta_l = \sigma_l$, σ_l can be a single semi-orthogonal value factor $Q^T\hat{Q}$, $\sigma_1 \leq \cdots \leq \sigma_d$. We are supposed maximize $\cos\theta_1$ to decrease errors of the sub-spaces.

For the theorem, there is θ_l relent in α_i, ($i = 1, \cdots, d$) or $\|x_{k+1} - \overline{x}_k\|_2$. Furthermore, maximize $\cos\theta_1$ is to enable $\cos\alpha_2, ..., \cos\alpha_d$ minimum, or $\cos\alpha_1$ maximum. As $1 < \eta_2^1 < \cdots < \eta_d^1$, $\cos\alpha_1$ is the principal content to decide $\cos\theta_1$. Thereby, $\cos\beta$ is maximized and select d as the tiny. The processes make η_i^1 has not any relation in $\|x_{k+r} - \overline{x}_k\|_2$, excludes the effect that $\|x_{k+r} - \overline{x}_k\|_2$ brought to θ ($r = 1, 2, \cdots$), or gets $\cos\theta_1$ at the largest degree approximately. This method can be expressed:

Utilizing $\hat{X}$: $\hat{X}_e = \hat{X}D_0$, where $D_0 = diag\left(1/\|\hat{x}_1\|_2, \cdots, 1/\|\hat{x}_n\|_2\right)$, we get $\hat{X}_e\hat{X}_e^T = \hat{X}D\hat{X}^T$, $D = D_0^2$. We should between centralized model and the orthogonal abjection, the maximum angle cosine sum of wealth; it can achieve the following optimization issue. Optimization function is as follows:

$$E_{AOGE} = \sum_{i=1}^{n} \cos^2\beta_i \qquad (5)$$

As $\cos^2\beta_i = \hat{x}'_{ei}QQ^T\hat{x}_{ei}$

where $\hat{x}_{ei} = \dfrac{\hat{x}_i}{\|\hat{x}_i\|}$ satisfies $\hat{x}'_{ei}\hat{x}_{ei} = 1$, there is

$$E_{AOGE} = \sum_{i=1}^{n} \hat{x}'_{ei}QQ^T\hat{x}_{ei} = \sum_{i=1}^{n} e'\hat{X}_e^TQQ^T\hat{X}_e e_i$$
$$= Tr\left(\hat{X}_e^TQQ^T\hat{X}_e\right) = Tr\left(Q^T\hat{X}_e\hat{X}_e^TQ\right) \qquad (6)$$

Let $Q \in R^{D\times d}$ is a serious of formed base of lower dimensional space and $M = \hat{X}_e\hat{X}_e^T = \hat{X}D\hat{X}^T$, then we have:

$$\max Tr\left(Q^TMQ\right)$$

$$s.t.\ Q^TQ = I \qquad (7)$$

Q denotes first d relevant eigenvectors largest M eigenvalues. The last, $Y = Q^TX$.

A. GODR Algorithm Description

Step one. Criminalizing points $X = [x_1, x_2, \cdots, x_n] \in R^{D\times n}$, so centralizing $\hat{X} = X\left(I - \dfrac{1}{n}ee^T\right)$;

Step two. Uniting covariance matrix: $\hat{x}_i \leftarrow \hat{x}_i/norm(\hat{x}_i)$, $i = 1, 2, \cdots, n$;

Step three. Assessing M or bringing in eigen values, getting eigen values $\lambda_1 \geq \cdots \geq \lambda_D$ or relevant eigenvector U;

Step four. Finding Q that can be the 1st d amount. Computing $Y = Q^TX$.

IV. EXPERIMENTS AND ANALYSIS

TABLE II.   EFFECTIVE RATE OF *PCA* AND GODR

$times_{exp}$	Increasing one $data_d$		$times_{exp}$	Increasing one $data_\theta$	
	rGODR	r_{PCA}		rGODR	r_{PCA}
350	95%	87.67%	350	98.77%	99.77%
450	92.00%	85.00%	450	98.85%	99.97%
550	93.00%	87.60%	550	98.50%	99.85%
650	93.00%	85.33%	650	97.60%	99.77%

To obtain regular datasets, we train the datasets having the quantities function or plus a regular datasets. We derive these basses and the lower dimensioned data. Utilizing PCA subspace dimension reduction and can be got from GODR. After reducing dimension utilizing PCA or GODR is got through $dist_{PCA}\left(Q_{PCA}, \hat{Q}_{PCA}\right)$ or $dist_{AOGE}\left(Q_{AOGE}, \hat{Q}_{AOGE}\right)$. With the method of subspace distance dimension reduction is effectiveness. With this offer, we are given the same criterion. PCA and GODR efficiency comparison effects can be found in tab II.

Use face expressed data, respectively choose the 1200 pictures, there has gone experimental results: GODR PCA. Experimental results show that GODR can be most effective, look Fig. 2). Face expression identification to verify the sample (Fig. 3). For the experiments, noised sample in PCA has enormous influence, however there is almost no effect GODR.

To discuss this, just dimension reduction method to find the original image is set low dimensional embedding. In general, the relevant data of semantic content is directly from these parameters. It formed many image dimension reduction image clustering or the earliest
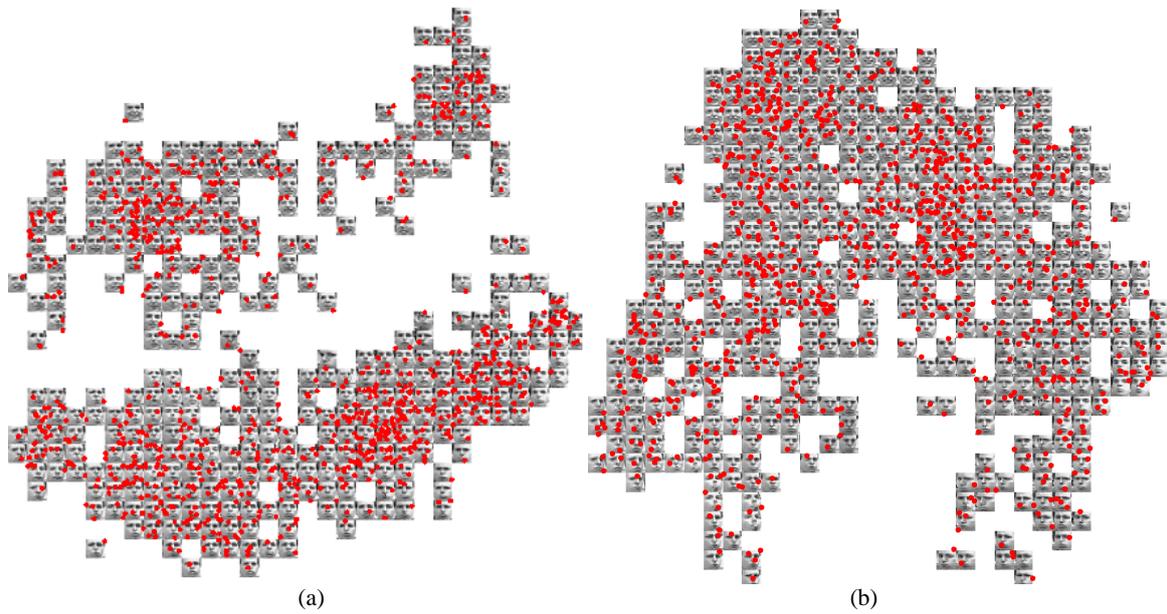
Figure 2.   Face expression identification, dimension from 650-3 ((a) dimension reduction through GODR (b) dimension reduction through PCA)
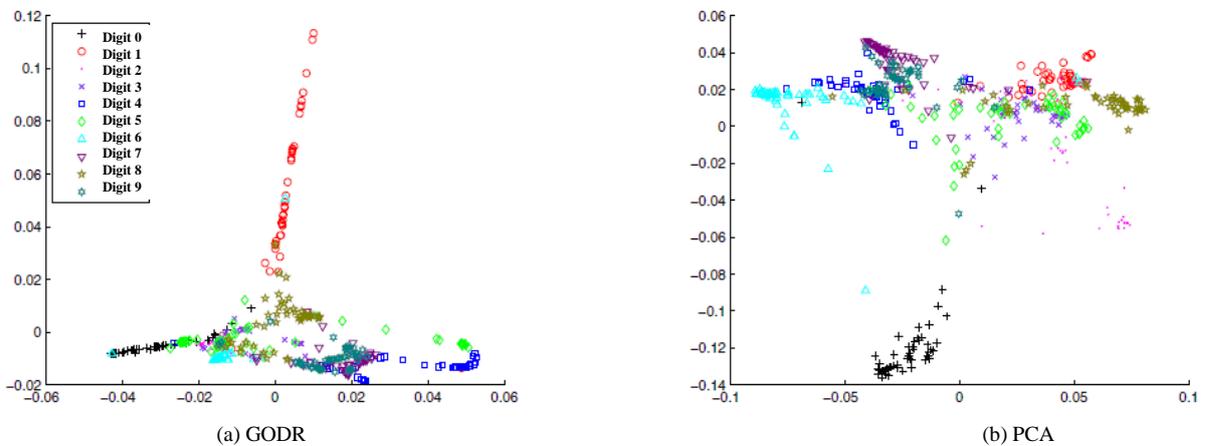

Figure 3.   Performed comparison: (a) GODR and (b) PCA.

application of visualization, and understands its parameters. In this section, we summarize the work methods and nonlinear dimension reduction as a first step, in order to facilitate the commonly used computer vision tasks, i.e. pose estimation, image/video segmentation, object tracking [27-30].

Fig. 3 compares the PCA method with GODR. GODR cluster on the vision, the query results of algorithm reveals the different types of Numbers, more compact, less than PCA for PCA exhibits a wider range [31,32] of dispersion and two dimensional embedded in regards to space objects. In addition, by each other in our results on the cluster analysis, the distribution of widely accepted belief of different structure, again to verify these figures. Cluster number between 0 and 1, for example, a far cry from, from other Numbers. The confirmed number of 0 and 1 discriminatory best precision is a widely accepted belief. In addition, due to the organization chart 9 and 7, 6 and 8 involves a common feature, the daughter's problem is a visualization in 2D space, can be concluded, from the embedded greeting.

V.   CONCLUSIONS

In this article, we present a novel global optimized dimensionality reduction (GODR) algorithm. This new algorithm to measure the relationship between the sample and the orthogonal projection data set, achieving global embedding dimension reduction based on the most optimal angle. GODR is a global linear methods, and there is a good degree of public and PCA. It uses the angle between the sample is measured; it is not the main component of sensitive data. GODR carry eigen decomposition in concentrated sample, to avoid overflowing the characteristics calculated using principal component analysis is very high dimensional samples. Facial expression recognition experiments confirmed, GODR have better recognition results than the PCA.

REFERENCES

[1]   Bellman, R.: Adaptive Control Processes: A Guided Tour, Princeton University Press, vol. 7, no. 4, pp. 225–228, 1961.

[2] Pless, R. and Simon, I.: Using thousands of images of an object, International Conference on Computer Vision, Pattern Recognition and Image Processing, vol. 16, no. 2, pp. 145–153, 2002.

[3] Pless, R.: Images Spaces and Video Trajectories: Using Isomap to Explore Video Sequences, Proc. IEEE Conference on Computer Vision and Pattern Recognition, vol. 17, no. 7, pp. 641–651, 2003.

[4] Brun, A.: Manifolds in Image Science and Visualization, PhD Thesis, Link oping University, vol. 9, no. 4, pp. 642–645, 2008.

[5] Lin, T. and Zha, H.: Riemannian Dimensionality reduction, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.30, no.5, pp. 796–809 ,2008.

[6] Tenenbaum, J. B., de Silva, V. and Langford, J. C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, vol. 290, no. 5500, pp. 2319–2323, 2000.

[7] Roweis, S. T. and Saul, L. K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding, Science, vol. 290, no. 5500, pp.2323–2326, 2000.

[8] Jenkins, O. C. and Mataric, M. J.: A spatio-temporal extension to Isomap nonlinear dimension reduction, ICML'04: Twenty-first international conference on Machine learning, New York, NY, USA, ACM Press, vol. 17, no. 3, pp. 437–441, 2004.

[9] Donoho, D. and Grimes, C.: When Does ISOMAP Recover the Natural Parameterization of Families of Articulated Images?, Technical report, Stanford University, vol. 14, no. 3, pp. 666–670, 2002.

[10] Zha, H. and Zhang, Z.: Isometric Embedding and Continuum ISOMAP, Proc. Of the Twentieth International Conference on Machine Learning (ICML-2003), Washington, DC , vol. 25, no. 4, pp. 734–750, 2003.

[11] De Silva, V. and Tenenbaum, J. B.: Global Versus Local Methods in Nonlinear Dimensionality Reduction, Advances in Neural Information Processing Systems (NIPS), S. Becker, S.T. and Obermayer, K. (Eds.), MIT Press, Cambridge, MA, pp.705–712, 2003.

[12] Brand, M.: Charting a Manifold, Advances in Neural Information Processing Systems (NIPS), S. Becker, S. T. and Obermayer, K. (Eds.), pp.961–968, 2003.

[13] Verbeek, J. J., Roweis, S. T. and Vlassis, N.: Non-linear CCA and PCA by Alignment of Local Models, Advances in Neural Information Processing Systems 16,Thrun, S., Saul, L. and Scholkopf, B. (Eds.), MIT Press, Cambridge, MA , pp. 34–37, 2004.

[14] Belkin, M. and Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering, Advances in Neural Information Processing Systems, pp. 164–177, 2002.

[15] Gerber, S., Tasdizen, T. and Whitaker, R.: Robust non-linear dimensionality reduction using successive 1-dimensional Laplacian Eigenmaps, ICML'07: Proc. 24th international conference on Machine learning, New York, NY, USA, ACM, pp.281–288, 2007.

[16] Brand, M.: Continuous Nonlinear Dimensionality Reduction by Kernel Eigenmaps, International Joint Conference on Artificial Intelligence, vol. 13, no. 9, pp. 437–441, 2003.

[17] Donoho, D. L. and Grimes, C.: Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data, PNAS, vol.100, no.10, pp.5591–5596, 2003.

[18] He, X. and Niyogi, P.: Locality Preserving Projections, Advances in Neural Information Processing Systems 16, Thrun, S., Saul, L. and Scholkopf, B. (Eds.), MIT Press, Cambridge, vol. 22, no. 2, pp. 332–337, 2004.

[19] De Ridder, D., Loog, M. and Reinders, M.: Local Fisher Embedding, Proc. International Conference on Pattern Recognition, pp.295–298, 2004.

[20] de Ridder, D., Kouropteva, O., Okun, O., Pietikainen, M. and Duin, R.: Supervised locally linear embedding, Artificial Neural Networks and Neural Information Processing, ICANN/ICONIP 2003 Proceedings, Lecture Notes in Computer Science 2714, Springer, vol. 32, no. 5, pp. 361–372, 2004.

[21] Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F. and Zucker, S.W.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, PNAS, vol.102, no.21, pp.7426–7431, 2005.

[22] Bengio, Y. and Monperrus, M.: Non-Local Manifold Tangent Learning, Advances in Neural Information Processing Systems , vol. 62, no. 6, pp. 69–79, 2005.

[23] Carreira-Perpin'an, M. A. and Zemel, R. S.: Proximity Graphs for Clustering and Dimensionality reduction, Advances in Neural Information Processing Systems 17, Saul, L.K., Weiss, Y. and Bottou, L. (Eds.), pp.225–232, 2005

[24] Weinberger, K. Q. and Saul, L. K.: Unsupervised Learning of Image Manifolds by Semidefinite Programming, Proc. IEEE Conference on Computer Vision and Pattern Recognition, Washington D.C., pp.988–995, 2004.

[25] Hinton, G. and Roweis, S.: Stochastic Neighbor Embedding, Advances in Neural Information Processing Systems (NIPS), vol. 96, no. 4, pp. 238–250, 2002

[26] Park, J., Zhang, Z., Zha, H. and Kasturi, R.: Local Smoothing for Dimensionality reduction, Proc. IEEE Conference on Computer Vision and Pattern Recognition, vol. 18, no. 7, pp. 452–459 ,2004.

[27] Verbeek, J. J., Roweis, S. T. and Vlassis, N.: Non-linear CCA and PCA by Alignment of Local Models, Advances in Neural Information Processing Systems 16,Thrun, S., Saul, L. and Scholkopf, B. (Eds.), MIT Press, Cambridge, MA , pp. 34–37, 2004.

[28] Belkin, M. and Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering, Advances in Neural Information Processing Systems, pp. 164–177, 2002.

[29] Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F. and Zucker, S.W.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, PNAS, vol.102, no.21, pp.7426–7431, 2005.

[30] Bengio, Y. and Monperrus, M.: Non-Local Manifold Tangent Learning, Advances in Neural Information Processing Systems , vol. 62, no. 6, pp. 69–79, 2005.

[31] Roweis, S. T. and Saul, L. K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding, Science, vol. 290, no. 5500, pp.2323–2326, 2000.

[32] Jenkins, O. C. and Mataric, M. J.: A spatio-temporal extension to Isomap nonlinear dimension reduction, ICML'04: Twenty-first international conference on Machine learning, New York, NY, USA, ACM Press, vol. 17, no. 3, pp. 437–441, 2004.

Xinbo Lan was born in Guangdong province of China in 1981. He has received his bachelor degree and master degree respectively. Currently, he is a lecturer. His research interests include Computer Network Security, Pattern Recognition and Embedding control.

Fault-Tolerant Transmission Protocol for Distant Agricultural Image Acquisition

Jian Chen, Deqin Xiao, Dongmin Liu, and Xiaoqing Jiang

College of Informatics, South China Agricultural University, Guangzhou 510642, China

Email: c-j1987@163.com, deqinx@scau.edu.cn, liudomy@stu.scau.edu.cn, jxq942@scau.edu.cn

Abstract—To solve the problem of the high cost in the GPRS communication and the limit transmission distance of WiFi, a transmission scheme for distant agriculture image acquisition was designed based on digital transmission radio in this paper. However, the majority of current digital transmission radio was designed for a small amount of data transmission. It could get a greater transmission distance with the help of the digital transmission radio, but the signal interference increased heavily when the digital transmission radio was used for image transmission. A fault-tolerant transmission protocol for agriculture image (FTTP-AI) based on digital transmission radio was designed in this paper. Packet verification was used to reduce the data errors caused by the signal interference of the digital transmission radio. At the same time, overtime retransmission and the lost packet retransmission were used to overcome the problem of packet loss. Experiments showed that the FTTP-AI could send the agriculture image to a remote computer center successfully in the field. With the help of the FTTP-AI, the rate of accuracy of data transmission was up to 99.2%, the success rate of image transmission was up to 95.8%, the costless distant transmission can achieve several kilometers. This scheme could satisfy the requirement of the low-cost for distant agriculture images transmission reliably.

Index Terms—Image Transmission, Fault-Tolerant Transmission Protocol, Packet Verification, Retransmission, Digital Transmission Radio

I. INTRODUCTION

The acquisition technique of the fast agricultural information is one of the key techniques in the field of the precise agriculture. The images teletransmission technique, as one kind of the information acquisition techniques, can be used in the agricultural field to perform tele-observation on the growing situation, plant diseases and insect pests of the crops [1]. In this way, the precise information collection of the site, as well as the real time monitoring of the field can be achieved to create the conditions for the modernization of the agricultural management. Nowadays, Wi-Fi, Blue Teeth, Zig-Bee and RFID[2, 3] are the main techniques in the short distance data wireless transmission, on the other hand, GPRS and 3G net are the main techniques in the long distance transmission [4, 5]. Xiong et al. [6] and Liu et al. [7] have already realized the above techniques respectively.

In the acquisition techniques of the distant agricultural images information, common Wi-Fi technique is unable to finish the long distance transmission [8]; however, GPRS [9, 10] and 3G [11] transmission depend on mobile business operator thus high expense cost is inevitable. It has become a bottleneck that the collection of images data several kilometers away frequently through the sensor net of the distant agricultural images. To solve the high cost of the GPRS as well as the limit transmission distance problem, the wireless digital transmission radio is used.

The DTR (digital transmission radio) [12] is a professional data transmission radio station of high-performance based on DSP and wireless radio techniques. The DTR is able to transmit various kinds of data such as digital voice, dynamic images, tele-control and tele-metering data etc. The DTR offers transparent RS232 interface and is able to achieve 19.2Kbps transmission rate as well as 2-10 kilometers transmission distance. In our earlier work [13], we used the DTR JZ875 to perform the transmission of the water data and achieved satisfied result of 6.5 kilometers transmission distance. Based on the above research and experiments, we studied the use of the DTR for the image transmission in this paper.

The commercial DTR JZ875 was used as the transceiver. In the transmission performance test of JZ875, bit error rate was found to be relative low and could be adopted to transmit images. However, JZ875 was interfered seriously due to the rapid increase of the amount of the images data in the process of transmission. This case brought problems such as the reception of the error data and the lost of the data packet.

In order to apply the DTR on the images transmission without above mentioned problems, a low-cost wireless sensor network for distant agriculture image acquisition was built in this paper. FTTP-AI (fault-tolerant transmission protocol for agricultural image) was specially designed to improve the rate of accuracy and success during the images transmission. In the rest of this paper, we would introduce our work in detail with the following structure. In part II, the system architecture for images transmission based on the DTR would be specified. In part III, we propose the protocol rule, the format of the packet and transmission procedure of FTTP-AI based on DTR. In part IV, the detail design of FTTP-AI such as Packet encapsulation and confirmation mechanism would be defined. In part V, the experimental results were given with the comparison of the data transmission effect with and without FTTP-AI and then

the statistic data would be presented to analysis these results. At the end of this paper, we conclude the achievements of our work as well as the standing problems in part VI.

II. THE SYSTEM ARCHITECTURE FOR IMAGES TRANSMISSION BASED ON THE DTR

Fig. 1 shows the system architecture for the agricultural images sensor network in this paper. Within the transmission distance of Wi-Fi, the images data collected by the images collection nodes was sent to the base station via Wi-Fi. Then within the transmission distance of the DTR, the images data collected by the base station was sent to the remote computer center via the DTR. In the next step, the remote computer center would send the images data to the system sever via 3G net or internet. After all, the user terminal could get the agricultural images from the system sever via the internet and perform further study on the images data.

In the actual deployment of the experiment, the base station in the Fig. 1 located at the experimental plots of Chen-cun, Tian-he district in Guangzhou, and the remote computer center located at the informatics college building of SCAU (South China Agricultural University). The linear distance between them is about 2.1 kilometers, which is out of the transmission distance of Wi-Fi. In order to cut down the cost, the commercial DTR JZ875 was used to transmit the images data in a long distance in this paper.
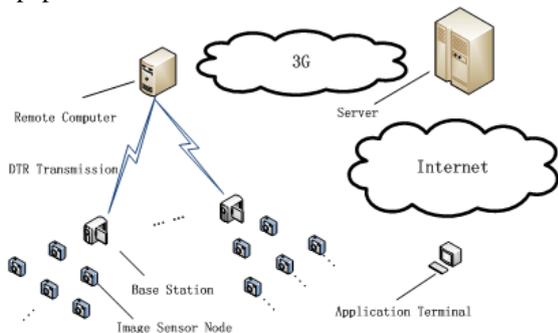
Figure 1.   System architecture

III. FTTP-AI BASED ON THE DTR

To solve the above mentioned problems in data transmission of the DTR, FTTP-AI based on the DTR was designed as follows.

A. The Design of the Fault-Tolerant Rule

One of the problems in the DTR transmission is the signal interference. Robust frame synchronization [14-17] was used to test the data in this paper, including bytes-count law and head-tail-mark law.

The other problem is the lost of the data packets in the DTR transmission. The continuum re-transmission mechanism [18] was used to check if there were lost data packets after all data was sent, and re-transmit the lost data packets.

B. The Design of the Packet

The unify format of the packet as tableI was adopted in this paper and there were eleven different type codes as tableII.

TABLE I.   THE FORMAT OF THE PACKET

Preamble	Type code	Packet data	Postamble

TABLE II.   TYPE SIZES FOR CAMERA-READY PAPERS

ID	Code	Size	Usage
1	0x01	12	SYN
2	0x02	11	SYN confirmation
3	0x03	512	Image data
4	0x04	10	Send completed
5	0x05	11	Send completed confirmation
6	0x06	10+count	Re-transmission list
7	0x07	11	Re-transmission list confirmation
8	0x08	512	Re-transmission of image data
9	0x09	10	Re-transmission completed
10	0x10	11	Re-transmission completed confirmation
11	0x11	13	Transmission confirmation

C. The Procedure of the Protocol

The procedure of the protocol was as Fig. 2. The sending terminal S and the receiving terminal R completed the images transmission base on the protocol.
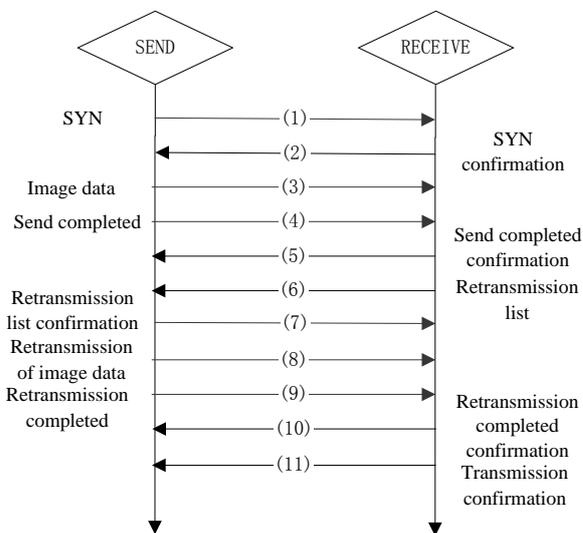

Figure 2.   The flowchart of the protocol

Step 1: The sending terminal S sends the synchronization data packets.

Step 2: The receiving terminal R returns the confirmation packets.

Step 3: The sending terminal S sends the images data packets.

Step 4: The sending terminal S sends the finished-mark packets.

Step 5: The receiving terminal R sends the completed confirmation packet to confirm if the sending is finished and check whether the images data received is integrity. If it is, then turn to Step 11; otherwise, go on Step 6.

Step 6: The receiving terminal R sends the packet including the list of the lost packets.

Step 7: The sending terminal S returns the packet to confirm the lost packets.

Step 8: The sending terminal S re-sends the lost packets.

Step 9: The sending terminal S re-sends the finished-mark packets.

Step 10: The receiving terminal R re-sends the completed confirmation packet to confirm if the sending is finished and check whether the images data received is integrity again. If it is, then go on Step 11; otherwise, turn to Step 6.

Step 11: The receiving terminal R sends transmission confirmation packet to confirm that the transmission is finished.

IV. THE DETAIL DESIGN OF THE FTTP-AI

A. Packet Encapsulation

In order to solve the problem of the data interference in the DTR transmission, the SYN code and bit-count were imported to allow for error control. The packet data was designed to include the preamble, initial length of the packet L_{packet}, initial data and the postamble. Four bytes EBH 90H EBH 90H were selected as the preamble, and three bytes EBH 90H EBH were selected as postamble. L_{packet} was included in the head of the packet to perform error control judgment.

When the receiving terminal received the data, the preamble and postamble would be decided whether it was right, if not, the data would be judged as the interference data and not to be processed in the next step; otherwise the packet would be unpacked, then the preamble and postamble would be extracted. Finally we would have L_{packet} from the head of the packet and judge if it was equal to $L_{received}$. If it was, the data would be accepted.

B. Confirmation Mechanism

Due to the possibilities of the lost packet in the transmission, the response mechanism was used to make sure that the packets could be sent to the receiving terminal successfully.

The total time of the packet transmitting from the sending terminal S to the receiving terminal R [19] is as:

$$t_{total} = t_{send} + t_{spread} + t_{receive} + t_{process} \qquad (1)$$

where t_{send} is the send-delay, t_{spread} is the transmit-delay, $t_{receive}$ is the receive-delay, $t_{process}$ is the process-delay. Among the four delay time, t_{spread} is decided by the velocity and the distance of the spread. The spread velocity of the radio waves in the air is 3*108m/s, and the spread distance of the DTR is 2-10 kilometers. So t_{spread} would be microsecond time and could be ignored here. Also, $t_{receive}$ is always ignored in the serial wireless communication [20] and so as the $t_{process}$ because the frequency of the ARM processor has already reached more than 100MHZ and the size of the packet as table Ⅰ is small enough. After all, it turns out to be as:

$$t_{total} = t_{send} = l / r \qquad (2)$$

where l is the length of the packet, r is the baud rate for serial ports. According to table Ⅰ, the max length of the packet is 512 bytes and the lowest spread velocity of the DTR JZ875 is 19200bit/s. So the maximum of t_{total} would be 512*8/19200=0.213s. Considering the round-send-delay and others kinds of delay, the overtime $t_{timeout}$ would be 1s in this paper.

The retransmission time out mechanism was designed as follows. When a packet which needed to be confirmed was sent out, a timer would be invoked. If the confirmation packet was received within $t_{timeout}$, the timer would stop and went on with next moves. If not, the packet would be judged as lost and retransmitted.

The integrity check mechanism for images data was designed as follows. When the packet with the type code 0x01 was received, the number of the image packets N_{packet} would be read out. $N_{received}$ would be compared with N_{packet} after the packet with the type code 0x05 or 0x10 was received. If they were equal then the packet with type code 0x11 would be returned and the image data would be saved, otherwise a list containing the lost packets would be returned instead.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimenalt Results without FTTP-AI

The data is easy to be interfered when transmitted using the DTR JZ875, which leads to error codes and lost packets. To detecting how the data was interfered, experiments have been conducted as follows: the test data with fixed size was sent after a fixed time interval, which would be analyzed after received.

a. The test result of the "100KB.txt"

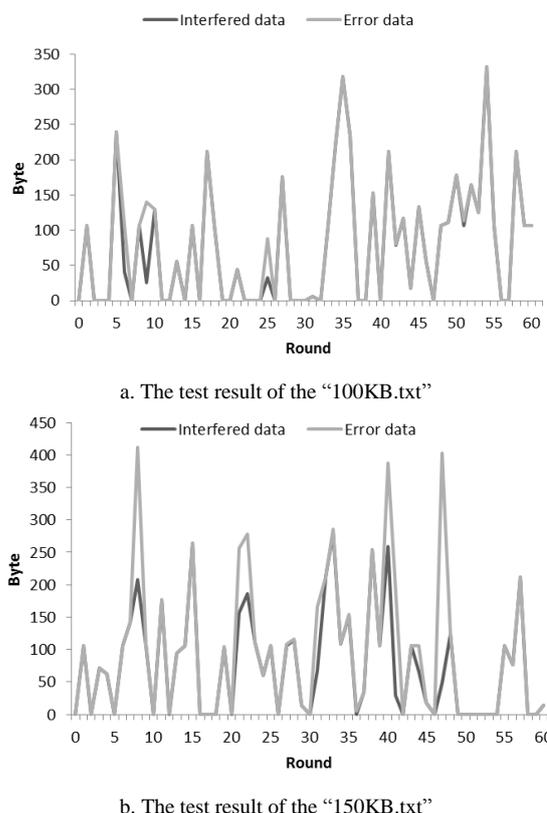
b. The test result of the "150KB.txt"

Figure 3.   Data analysis without any protocol

As Fig. 1, the cameras with 1.3 million pixels were adopted as the images collection nodes. The size of the photos collected was between 100KB and 150KB. So the 100KB and 150KB were set to be the threshold to check the size of the files. The test files "100KB.txt" and "150KB.txt" contained all '1' text and all '0' text respectively. The sending terminal took turns to send "100KB.txt" and "150KB.txt" with one hour interval. The total time of the test was 120 hours. The result of the test was as Fig. 3.

It can be seen in the Fig. 3 that the data was interfered seriously during the transmission without any protocols, and the interfered data even changed without rule. Compared with Fig. 3a, obviously Fig. 3b shows more serious interference. It follows that the larger the amount of data is, the more serious the interference would be. The amount of the interfered data was equal to that of the error data in most of the test periods. But in a few cases the amount of the error data was larger than the interfered, which was caused by the crosstalk-induced effects on the normal receiving data.

B. Experimental Results with FTTP-AI

In the following experiments, the FTTP-AI was applied in the transmission with the DTR JZ875.
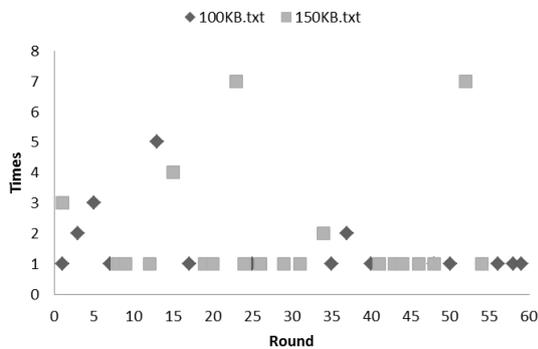
(a) The statistics of the re-transmission
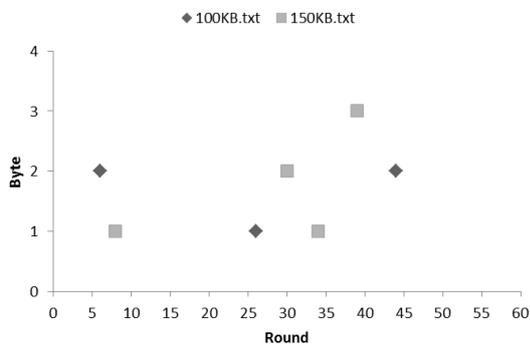
(b) The statistics of the error code

Figure 4.   Data analysis with FTTP-AI

In the text files transmission experiment, the test files were the same as above mentioned in the experiments without FTTP-AI. "100KB.txt" and "150KB.txt" contained all '1' text and all '0' text respectively. The sending terminal took turns to send "100KB.txt" and "150KB.txt" with one hour interval. The total time of the test was 120 hours. FTTP-AI was applied to the

transmission between the sending terminal and the receiving terminal.

The experimental result shows that the amount of the sending data is equal to that of the receiving data, which means the interference data is zero. Fig. 4 shows the number of the re-transmission of the sending terminal as well as the error rate of the receiving terminal.

It can be seen from the Fig. 4a that the rate of re-transmission tends to be larger on the large files more than the small files. In most cases the data could be transferred integrally through re-transmission just one time even when it was interfered in the DTR transmission, and only in a few cases several times of re-transmission were needed.

It can be seen from the Fig. 4b that there are few cases where error codes occur. Error codes happened to occur in 7 of 120 files. The total error codes were 12 bytes. So the rate of success to transmit text files is (120-7)/120*100%=94.2%, which means the total rate of the error codes is 12/ (100*60+150*60)*100%=0.08%. It is effective to improve the text files transmission rate of success using the FTTP-AI.

In the image files transmission experiment, the real-time images collected from the farmland were transmitted to the remote computer center. The linear transmission distance was 2.1 kilometers and the sending interval was one hour. The total time of the test was 120 hours. FTTP-AI was applied to the transmission between the sending terminal and the receiving terminal.

The experimental result shows that 120 image files with size between 94KB and 124KB were sent, and the receiving terminal received all the image files. Fig. 5 shows the number of re-transmission in the image files transmission experiment. Most of the image files could be transferred integrally after 1 to 3 times re-transmission. Few periodic images should be re-transmitted more than 3 times.
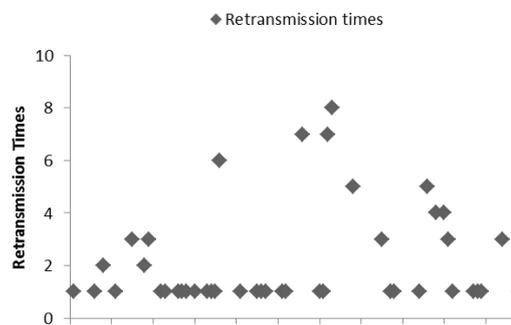

Figure 5.   Retransmission times of image transmission

In the transmission test of the image files, there were 5 images which error code occurred. Compared with the effect of the image received successfully (on the left), Fig. 6 shows the example of image received with error code (on the right). So the rate of success to transmit image files was (120-5)/120*100%=95.8%. It is effective to improve the image files transmission rate of success using the FTTP-AI.

Figure 6.   Contrast for image transmission

VI.   CONCLUSION

To solve the problem of the high cost of the GPRS and the limit transmission distance of the Wi-Fi, the DTR was imported to the transmission of the agricultural images and the FTTP-AI protocol was designed based on the DTR in this paper. The study offered a solution to cover the low-cost requirement, at the same time insured the distant transmission of the agricultural images. With FTTP-AI, the rate of accuracy of data transmission was up to 99.2%, the success rate of image transmission was up to 95.8%, the costless distant transmission could achieve several kilometers. This scheme could satisfy the requirement of the low-cost for distant agriculture images transmission reliably.

Even though the application of the FTTP-AI on the DTR JZ875 can realize the low-cost for distant agriculture images transmission, which is perfect for agriculture because of the relative low data amount requirement. However, for transmission with much further distance, the rate of error codes will increase rapidly due to the signal attenuation of the DTR. Meanwhile, the FTTP-AI protocol is not fast enough for the situation where the images data need to be transmitted frequently for the data amount limit of the DTR itself. To perform transmission with further distance, a proposal integrated with high efficiency correcting error mechanism and 3G would be the key points in the future study.

REFERENCES

[1] Z. Xiaochao, W. Yiming, F. Xianfa and Z. Huaping. Information Acquisition Techniques of Precision Agriculture. Transactions of Chinese Society of Agricultural Machinery, vol. 6, 2002, pp. 125-128.

[2] M. Miura, S. Ito, R. Takatsuka, T. Sugihara, and S. Kunifuji, An empirical study of an rfid mat sensor system in a group home, *Journal of Networks*, vol. 4, (no. 2), 2009, pp. 133-139.

[3] H. Liu, Y. Chen and Y. Chen, A frequency diverse Gen2 RFID system with isolated continuous wave emitters, *Journal of Networks*, vol. 2, (no. 5) , 2007, pp. 54-60.

[4] S. Yuwen, S. Mingxia, Z. Xiangfu, X. Yingjun and Z. Liang. Design of Embedded Agricultural Intelligence Services System Based on ZigBee Technology. Transactions of the Chinese Society for Agricultural Machinery, vol. 5, 2010, pp. 029.

[5] S. Hengstler and H. Aghajan, WiSNAP: a wireless image sensor network application platform. IEEE, 2006, pp. 6 -12.

[6] X. Yinjin, S. Mingxia, S. Yuwen, X.You and L. Xiangze. Design on System of Acquisition and Wireless Transmission for Farmland Image. Transactions of the Chinese Society for Agricultural Machinery, vol. 3, 2011, pp. 037.

[7] L. Longshen, S. Mingxia, S. Yuwen, L. Mingzhou and X. Yinjun. Acquisition System and Wireless Transmission by 3G for Farmland Image Based on FPGA. Transactions of the Chinese Society for Agricultural Machinery, vol. 12, 2011, pp.186-190.

[8] I.C.S.L. Committee, Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. IEEE Std., 1997.

[9] J.M. Oliveira and E. Carrapatoso, Dynamic Generation of SMIL-Based Multimedia Interfaces, *Journal of Multimedia*, vol. 3, (no. 4), 2008, pp. 14-25.

[10] T. Pliakas, G. Kormentzas and C. Skianis, Scalable video streaming traffic delivery in IP/UMTS networking environments, *Journal of Multimedia*, vol. 2, (no. 2) , 2007, pp. 37-46.

[11] M. Li, Z. Xu and Y. Xu, A Robust Video Coding Scheme for Mobile Wireless Communications, *Journal of Multimedia*, vol. 4, (no. 6) , 2009, pp. 405-411.

[12] R. Li. Application of data set station in computar supervision. Electrical Drive Automation, vol. 5, 2004, pp.48-50.

[13] X. Deqin, G. Zhichun, F. Jianzhao, X. Kehui, and L. Xiwen, Design and experiment of wireless sensor networks for paddyfield moisture monitoring. Transactions of the Chinese Society of Agricultural Engineering, vol. 2, 2011, pp. 030.

[14] Massey J. Optimum frame synchronization. Communications, IEEE Transactions on, vol. 2, 1972, pp.115-119.

[15] Steinga A, Wijngaarden A J, Teich W. Frame synchronization using superimposed sequences. Proc. IEEE ISIT., 1997, pp. 489.

[16] Dodds D, Button L, Pan S M. Robust frame synchronization for noisy PCM systems. Communications, IEEE Transactions on. vol. 5, 1985, pp.465-469.

[17] Christodoulides L M, Braithwaite S J, Chandler S. Frame synchronisation using two frame alignment words. Electronics Letters, vol. 2, 1994, pp.104-106.

[18] De Turck K, Wittevrongel S. Delay analysis of the Go-Back-N ARQ protocol over a time-varying channel. Formal Techniques for Computer Systems and Business Processes. 2005, pp.124-138.

[19] M Maróti, B Kusy, G Simon, Á Lédeczi. The flooding time synchronization protocol. Proceedings of the 2nd international conference on Embedded networked sensor systems ACM, 2004, pp.39-49.

[20] Y. Zongkai, Z. Dasheng, W. Yuming, C. Wenqing and H. Jianhua. Survey of time synchronization algorithms in wireless sensor network. Computer Applications, vol. 5, 2005, pp. 1170-1172.

Jian Chen received his B.Sc. degree in Computer Science and Technology in 2010 from South China Agricultural University,

Guangzhou, China. He is now a M.S. candidate in the college of Informatics, South China Agricultural University, Guangzhou, China. His research interest includes computer network, Agricultural Engineering.

Deqin Xiao received her Ph.D. degree in Agricultural Electrification and Automation in 2009 from South China Agricultural University, Guangzhou, China. She is now a professor in the college of Informatics, South China Agricultural University, Guangzhou, China. Her research interest includes wireless sensor network, Agricultural Engineering.

Dongmin Liu received her B.Sc. degree in Computer Science and Technology in 2010 from South China Agricultural University, Guangzhou, China. She is now a M.S. candidate in the college of Informatics, South China Agricultural University, Guangzhou, China. Her research interest includes bioinformatics, Agricultural Engineering.

Xiaoqing Jiang received his B.Sc. degree in Communication Engineering in 2005 from South China Normal University, Guangzhou, China. He is now a M.S. candidate in the college of Informatics, South China Agricultural University, Guangzhou, China. His research interest includes computer network, wireless sensor network.

Call for Papers and Special Issue Proposals

Aims and Scope.

Journal of Multimedia (JMM, ISSN 1796-2048) is a scholarly peer-reviewed international scientific journal published bimonthly, focusing on theories, methods, algorithms, and applications in multimedia. It provides a high profile, leading edge forum for academic researchers, industrial professionals, engineers, consultants, managers, educators and policy makers working in the field to contribute and disseminate innovative new work on multimedia.

The Journal of Multimedia covers the breadth of research in multimedia technology and applications. JMM invites original, previously unpublished, research, survey and tutorial papers, plus case studies and short research notes, on both applied and theoretical aspects of multimedia. These areas include, but are not limited to, the following topics:

- Multimedia Signal Processing
- Multimedia Content Understanding
- Multimedia Interface and Interaction
- Multimedia Databases and File Systems
- Multimedia Communication and Networking
- Multimedia Systems and Devices
- Multimedia Applications

JMM EDICS (Editors Information Classification Scheme) can be found at http://www.academypublisher.com/jmm/jmmedics.html.

Special Issue Guidelines

Special issues feature specifically aimed and targeted topics of interest contributed by authors responding to a particular Call for Papers or by invitation, edited by guest editor(s). We encourage you to submit proposals for creating special issues in areas that are of interest to the Journal. Preference will be given to proposals that cover some unique aspect of the technology and ones that include subjects that are timely and useful to the readers of the Journal. A Special Issue is typically made of 10 to 15 papers, with each paper 8 to 12 pages of length.

The following information should be included as part of the proposal:

- Proposed title for the Special Issue
- Description of the topic area to be focused upon and justification
- Review process for the selection and rejection of papers.
- Name, contact, position, affiliation, and biography of the Guest Editor(s)
- List of potential reviewers
- Potential authors to the issue
- Tentative time-table for the call for papers and reviews

If a proposal is accepted, the guest editor will be responsible for:

- Preparing the "Call for Papers" to be included on the Journal's Web site.
- Distribution of the Call for Papers broadly to various mailing lists and sites.
- Getting submissions, arranging review process, making decisions, and carrying out all correspondence with the authors. Authors should be informed the Instructions for Authors.
- Providing us the completed and approved final versions of the papers formatted in the Journal's style, together with all authors' contact information.
- Writing a one- or two-page introductory editorial to be published in the Special Issue.

Special Issue for a Conference/Workshop

A special issue for a Conference/Workshop is usually released in association with the committee members of the Conference/Workshop like general chairs and/or program chairs who are appointed as the Guest Editors of the Special Issue. Special Issue for a Conference/Workshop is typically made of 10 to 15 papers, with each paper 8 to 12 pages of length.

Guest Editors are involved in the following steps in guest-editing a Special Issue based on a Conference/Workshop:

- Selecting a Title for the Special Issue, e.g. "Special Issue: Selected Best Papers of XYZ Conference".
- Sending us a formal "Letter of Intent" for the Special Issue.
- Creating a "Call for Papers" for the Special Issue, posting it on the conference web site, and publicizing it to the conference attendees. Information about the Journal and Academy Publisher can be included in the Call for Papers.
- Establishing criteria for paper selection/rejections. The papers can be nominated based on multiple criteria, e.g. rank in review process plus the evaluation from the Session Chairs and the feedback from the Conference attendees.
- Selecting and inviting submissions, arranging review process, making decisions, and carrying out all correspondence with the authors. Authors should be informed the Author Instructions. Usually, the Proceedings manuscripts should be expanded and enhanced.
- Providing us the completed and approved final versions of the papers formatted in the Journal's style, together with all authors' contact information.
- Writing a one- or two-page introductory editorial to be published in the Special Issue.

More information is available on the web site at http://www.academypublisher.com/jmm/.

(Contents Continued from Back Cover)