Journal of Multimedia

Contents

An Important Frame Distinction Model of Stereoscopic Video Based on Content

Xiaodong Wang, Tengfei Wang, Binbin Hu, Gangyi Jiang, and Lianjun Zhang

Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, Zhejiang China

Email: wangxiaodong0505@gmail.com

Abstract—As different frame lost will lead to different distortion for stereoscopic video sequence, we take into account the temporal and spatial correlation of stereo video, error diffusion characteristics left and of right view frame, as well as recursion theory. Based on the above, we put forward a frame important distinction model of stereoscopic video based on content. The experimental results show that the model can be applied for stereoscopic video sequences with different motion intensity and disparity, accordingly, code ends can accurately estimate the importance of each frame for terminal perception. Finally, the applicability of the model is discussed.

Index Terms—Frame Importance; Content Perception; Error Diffusion; Spatial-Temporal Correlation

I. INTRODUCTION

In recent years, researches on technologies related with stereo video have achieved much attention along with the rapid development of multimedia [1-2]. As the amount of stereo video data is very huge, due to insufficient bandwidth, transmission delay and delay jitter during the video stream transmission process, quality of terminal perception will be reduced. In stereo video, as different frame and different content play different role on terminal perception, the sensitive level for loss is also different [3]. Hence, it is essential to focus on how to distinguish the importance of stereo video frame as to improve the transmission efficiency by carrying out differentiated services [4].

Currently, there are two ways to distinguish the frame importance. One way is based on the dependent relationship of the code to distinct the frame importance. In one Group of Picture (GOP), frame I is generally considered as the most important part, then followed by frame P and frame B [5]. Hartanto F and Sirisena H R. Hybrid (1999) suggest a way to determine the frame importance according to different frame type, and the importance of frame I and frame P is set as 1.15:0.35 [6]. Lin C H and other scholars propose importance model based on the position of frame and coding sequence of video data from the viewpoint of data frame error recovery, and frame I, frame B and frame P are divided into different levels to guide the data transmission [7-8]. Since the content characteristics of stereoscopic video frame are different, it is not reasonable to distinguish frame importance by position because the above methods ignore the relationship between video frames and the

effects of error diffusion. Yang et al (2009) take error diffusion into consideration and establish the ELEP model to measure the frame importance [9]. The model assumes that the earlier errors occur, more serious error diffuse, and the greater impact on the subsequent frames. Thus, the relationship between frame importance and its position in the GOP is approximately linear. However, the video sequence factor is not considered in the ELEP model, thus it fails to reflect the difference between high motion sequence and low motion sequence.

The second way is to distinguish video content that based on grammatical structure. The current H.264/AVC [10] divides the video data into three levels by encoded film to distinguish the importance. Scalable coding, such as fine scalable coding and progressive fine granularity scalable coding divide coding data into the base layer and enhancement layer [11-12], and subsequent enhancement layer must rely on the basic layer data decoding. So it is considered that the importance of basic layer is higher than the subsequent enhancement one. Though the method based on grammar structure based is simple, it is not accurate enough to make quantitative analysis.

The above analysis shows the impact of video content characteristics on the frame importance are not considered in the above two methods, therefore it can not accurately reflect the contribution of frame to the terminal perception, and they are only suitable for single-view video. Compared with the single-view video, stereo video asks for deeper and differentiated information, relying on higher spatial-temporal correlation and coding-decoding dependency, and the error diffusion relationship between left and right views is more complex. Therefore, it is more complex to distinguish the frame importance.

This paper aims to establish a frame importance model of stereoscopic video based on content-aware by taking into consideration of video content characteristics, spatial-temporal correlation, disparity information as well as network transmission and error concealment algorithm.

Section II analyzes measuring methods of the frame importance of stereo video, a frame importance model of stereo view is presented. In Section III, simulations are conducted to validate the frame importance model. Finally, paper is concluded in Section IV.

II. FRAME IMPORTANCE MODEL OF STEREO VIDEO

Since different video frame and content make different

contribution to the perception of the terminal video, stereo video frame importance is defined as follows in this paper: loss of one frame of stereo video will distort the entire sequence, and greater distortion shows that greater contribution to the perception made by the frame.

In this thesis, stereo video coding structure is IPPP, and the first frame is coded as frame I, while the remaining frames are coded as frame P, moreover, frame I is regarded as the most important frame. As stream of stereo video encodings have predictive relationship, loss of one frame not only distorts current frame, but also affects the subsequent frames. So important parameters of the stereo video s view t moment frames can be expressed as:

$$I(s,t) = I_c(s,t) + I_p(s,t) \qquad (1)$$

$I_c(s,t)$ refers to the error concealment distortion whose distortion is caused by a lost frame after error recovery. $I_p(s,t)$ refers to the error diffusion distortion as the loss of one frame may affect the subsequent frames. Because the frame importance measure is related with the specific frame error concealment methods, "frame copy" error concealment method is employed in this thesis for data recovery, which is to copy the previous frame of the same view when a frame of left view is lost. Due to special location, left view first frame will be copied when the first frame of right view is lost.

Stereo video has left and right view and different coding. During the process of encoding and decoding, right view frames will refer to left view ones while the left view frames will not refer to the right one, hence their error diffusion differ. In the following section, we will make research on frame importance model in left and right view.

A. Frame Importance Model of Right View

In light view, its previous frame will be copied when a frame is lost, therefore, the error concealment distortion can be seen as:

$$I_c(s,t) = D(s,t) \quad (s=1) \qquad (2)$$

where $D(s,t)$ represents mean square error (MSE) between frame of right view and left view at t time. Loss of right view frame only affects subsequent right view frames, thus its distortion is only related to inter-coded macroblocks (MB). Assuming the subsequent frame intra prediction ratio is β. When a frame lost, the effect to the subsequent λ frame:

$$D(s,t+\lambda) = (1-\beta) * \alpha * D(s,t+\lambda-1) \qquad (3)$$

the distortion of the subsequent frame due to loss of one frame can be seen as:

$$D(s,t+\lambda) = (1-\beta)^\lambda * \alpha^\lambda * D(s,t) \,(s=1) \qquad (4)$$

Parameter α is error diffusion factor which shows distortion ratio of next frame macroblocks participating prediction and present frame. α involves with sequential movement intensity and it is smaller than 1 but bigger than 0 in a video sequence which changes steadily, we

can assume $\alpha = (1-\beta)^2$. Therefore, the error diffusion distortion can be derived as:

$$
\begin{aligned}
I_p&(s,t) \\
&= D(s,t+1) + D(s,t+2) + ... + D(s,t+\lambda) \\
&= (1-\beta) * \alpha * D(s,t) + (1-\beta)^2 * \alpha^2 * D(s,t) \\
&\quad + ... + (1-\beta)^\lambda * \alpha^\lambda * D(s,t) \qquad (s=1) \\
&= \frac{(1-\beta) * \alpha * \left[1-(1-\beta)^\lambda * \alpha^\lambda\right]}{1-(1-\beta) * \alpha} * D(s,t) \\
&= \frac{\left[1-(1-\beta) * \alpha\right] - (1-\beta)^{\lambda+!} * \alpha^{\lambda+1}}{1-(1-\beta) * \alpha} * D(s,t)
\end{aligned}
$$
(5)

In conclusion, the expected frame importance in right view is given by:

$$
\begin{aligned}
I(s,t) &= I_c(s,t) + I_p(s,t) \\
&= D(\mathrm{s},\mathrm{t}) \\
&\quad + \frac{\left[1-(1-\beta) * \alpha\right] - (1-\beta)^{\lambda+!} * \alpha^{\lambda+1}}{1-(1-\beta) * \alpha} * D(s,t) \\
&= \left[\frac{\left[1-(1-\beta) * \alpha\right] - (1-\beta)^{\lambda+!} * \alpha^{\lambda+1}}{1-(1-\beta) * \alpha} + 1\right] \\
&\quad * D(s,t) \\
&= \frac{1-(1-\beta)^{\lambda+!} * \alpha^{\lambda+1}}{1-(1-\beta) * \alpha} * D(s,t)
\end{aligned}
$$
(6)

where $s=1$, $D(0,0)=0$.

B. Frame Importance Model of Left View

For the left video frames, error concealment distortion can be derived as follows:

$$I_c(s,t) = D(s,t) \,(s=0) \qquad (7)$$

Loss of a frame in left view will affect all subsequent frames, and the error diffusion can be denoted as:

$$I_P(s,t) = I_{PL}(s,t) + I_{PR}(s,t) \,(s=0) \qquad (8)$$

$I_{PL}(s,t)$ means the total distortion of subsequent left view frames due to the loss of left view frame , while $I_{PR}(s,t)$ is the total distortion of subsequent right view frames due to the loss of left view frame. Similar as the previous section, $I_{PL}(s,t)$ can be derived as:

$$
\begin{aligned}
I_{PL}&(s,t) = D(s,t+1) + D(s,t+2) + ... + D(s,t+\lambda) \\
&= (1-\beta) * \alpha * D(s,t) + (1-\beta)^2 * \alpha^2 * D(s,t) \\
&\quad + ... + (1-\beta)^\lambda * \alpha^\lambda * D(s,t) \qquad (s=0) \\
&= \frac{(1-\beta) * \alpha * \left[1-(1-\beta)^\lambda * \alpha^\lambda\right]}{1-(1-\beta) * \alpha} * D(s,t) \\
&= \frac{\left[1-(1-\beta) * \alpha\right] - (1-\beta)^{\lambda+!} * \alpha^{\lambda+1}}{1-(1-\beta) * \alpha} * D(s,t)
\end{aligned}
$$
(9)

Error diffusion distortion of the subsequent right view frames caused by the loss of left view frame can be

derived by weighing of previous view frames' distortion at present viewpoint and same view frames' distortion previously, so, error diffusion distortion for subsequent λ frame in right view can be expressed as:

$$
\begin{aligned}
&D(s+1,t+\lambda) \\
&= (1-\beta)*(1-\gamma)*\omega_1*D(s,t+\lambda) \\
&\quad +(1-\beta)*\gamma*\omega_2*D(s+1,t+\lambda-1) \qquad (s=0) \qquad (10)\\
&= a*D(s,t+\lambda)+b*D(s+1,t+\lambda-1)
\end{aligned}
$$

where $a=(1-\beta)*(1-\gamma)*\omega_1$, $b=(1-\beta)*\gamma*\omega_2$. γ is the proportion of intra-frame prediction, while Beta stands for inter prediction ratio, and ω_1 and ω_2 are weighting factors, and $\omega_1+\omega_2=1$. Therefore, total distortion for subsequent λ frame in right view can be expressed as:

$$
\begin{aligned}
&I_{PR}(s,t)=D(s+1,t)+D(s+1,t+1)\\
&+...+D(s+1,t+\lambda)=a*\begin{bmatrix}D(s,t)+D(s,t+1)\\+...+D(s,t+\lambda)\end{bmatrix}\\
&+a*b*\begin{bmatrix}D(s,t)+D(s,t+1)\\+...+D(s,t+\lambda-1)\end{bmatrix}\\
&+a*b^2*\begin{bmatrix}D(s,t)+D(s,t+1)\\+...+D(s,t+\lambda-2)\end{bmatrix}\\
&+a*b^3*\begin{bmatrix}D(s,t)+D(s,t+1)\\+...+D(s,t+\lambda-3)\end{bmatrix}\\
&+...++a*b^\lambda*D(s,t)\\
&=a*\frac{1-(1-\beta)^{\lambda+1}*\alpha^{\lambda+1}}{1-\alpha+\alpha*\beta}*D(s,t)\\
&+a*b*\frac{1-(1-\beta)^\lambda*\alpha^\lambda}{1-\alpha+\alpha*\beta}*D(s,t)\\
&+a*b^2*\frac{1-(1-\beta)^{\lambda-1}*\alpha^{\lambda-1}}{1-\alpha+\alpha*\beta}*D(s,t)\\
&+a*b^3*\frac{1-(1-\beta)^{\lambda-2}*\alpha^{\lambda-2}}{1-\alpha+\alpha*\beta}*D(s,t)\\
&+...++a*b^\lambda*D(s,t)
\end{aligned}
$$

$$
=\begin{bmatrix}a*\dfrac{1-(1-\beta)^{\lambda+1}*\alpha^{\lambda+1}}{1-\alpha+\alpha*\beta}\\[2ex]+a*b*\dfrac{1-(1-\beta)^\lambda*\alpha^\lambda}{1-\alpha+\alpha*\beta}\\[2ex]+a*b^2*\dfrac{1-(1-\beta)^{\lambda-1}*\alpha^{\lambda-1}}{1-\alpha+\alpha*\beta}\\[2ex]+a*b^3*\dfrac{1-(1-\beta)^{\lambda-2}*\alpha^{\lambda-2}}{1-\alpha+\alpha*\beta}\\[2ex]+...+a*b^\lambda\end{bmatrix}*D(s,t) \qquad (s=0) \quad (11)
$$

$I_P(s,t)$ can be expressed as:

$$
\begin{aligned}
&I_P(s,t)\\
&=I_{PL}(s,t)+I_{PR}(s,t)\\
&=\frac{[1-(1-\beta)*\alpha]-(1-\beta)^{\lambda+!}*\alpha^{\lambda+1}}{1-(1-\beta)*\alpha}*D(s,t)
\end{aligned}
$$

$$
+\begin{bmatrix}a*\dfrac{1-(1-\beta)^{\lambda+1}*\alpha^{\lambda+1}}{1-\alpha+\alpha*\beta}\\[2ex]+a*b*\dfrac{1-(1-\beta)^\lambda*\alpha^\lambda}{1-\alpha+\alpha*\beta}\\[2ex]+a*b^2*\dfrac{1-(1-\beta)^{\lambda-1}*\alpha^{\lambda-1}}{1-\alpha+\alpha*\beta}\\[2ex]+a*b^3*\dfrac{1-(1-\beta)^{\lambda-2}*\alpha^{\lambda-2}}{1-\alpha+\alpha*\beta}\\[2ex]+...+a*b^\lambda\end{bmatrix}*D(s,t)
$$

$$
=\begin{bmatrix}(a+1)*\dfrac{1-(1-\beta)^{\lambda+1}*\alpha^{\lambda+1}}{1-\alpha+\alpha*\beta}\\[2ex]+a*b*\dfrac{1-(1-\beta)^\lambda*\alpha^\lambda}{1-\alpha+\alpha*\beta}\\[2ex]+a*b^2*\dfrac{1-(1-\beta)^{\lambda-1}*\alpha^{\lambda-1}}{1-\alpha+\alpha*\beta}\\[2ex]+a*b^3*\dfrac{1-(1-\beta)^{\lambda-2}*\alpha^{\lambda-2}}{1-\alpha+\alpha*\beta}\\[2ex]+...+a*b^\lambda\end{bmatrix}*D(s,t) \qquad (s=0) \quad (12)
$$

In summary, the expected frame importance in left view can be indicated as:

$$
\begin{aligned}
&I(s,t)\\
&=I_c(s,t)+I_P(s,t)\\
&=I_c(s,t)+I_{PL}(s,t)+I_{PR}(s,t)
\end{aligned}
$$

$$
=\begin{bmatrix}1+\dfrac{(1-\beta)*\alpha-[(1-\beta)*\alpha]^{\lambda+1}}{1-*\alpha+\alpha*\beta}\\[2ex]+\dfrac{a*(1-b^{\lambda+1})}{(1-b)*(1-*\alpha+\alpha*\beta)}\\[2ex]+\dfrac{a*[(1-\beta)*\alpha]^{\lambda+2}}{(1-b)*(1-*\alpha+\alpha*\beta)}\\[2ex]-\dfrac{a*[(1-\beta)*\alpha]*b^{\lambda+1}}{(1-b)*(1-*\alpha+\alpha*\beta)}\end{bmatrix}*D(s,\mathrm{t}) \qquad (s=0) \quad (13)
$$

$$
=\begin{bmatrix}1+\dfrac{c-c^{\lambda+1}}{1-c}+\dfrac{a*(1-b^{\lambda+1})}{(1-b)*(1-c)}\\[2ex]+\dfrac{a*(c^{\lambda+2}-c*b^{\lambda+1})}{(1-b)*(1-c)}\end{bmatrix}*D(s,\mathrm{t})
$$

where λ refers to the number of subsequent frames, $a=(1-\beta)*(1-\gamma)*\omega_1$, $b=(1-\beta)*\gamma*\omega_2$, $c=(1-\beta)*\alpha$

III. SIMULATION RESULT

In this section, we present simulations to verify our analysis and evaluate the performance of our proposed frame importance model of stereoscopic video.

We select standard testing sequence for performance evaluation, i.e., a high motion sequence "crowd", and a low motion sequence "book". The frame rate is set to 30 frames per second (fps), QP takes the constant 28. Stereoscopic video are encoded as standard H.264/MVC on JM18.2 platform. The P-frame data is subject to random frame loss at rates of 10%. We make error recovery based on the aforesaid error concealment

method, and experimental results are shown in Figure I and Figure II. PSNR (Peak Signal to Noise Ratio) represents frame importance. we define stereo video frame importance as average PSNR of stereo video after loss of one frame, the importance of one frame is inversely proportional to model value, then we convert MSE to PSNR, which using the simple relation $PSNR = 10 * log_{10} \frac{255^2}{MSE}$, results shown in Figure III and Figure IV. In order to simulate the random packet loss in network, we take many different experiments in the same condition and calculate their average.
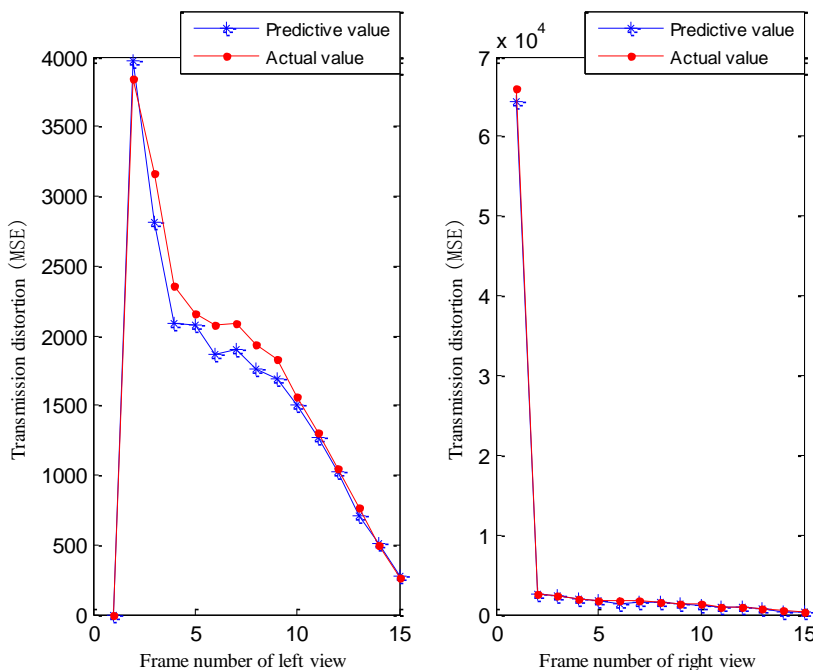
Figure 1. The total distortion of entire sequence due to loss of one crowd frame
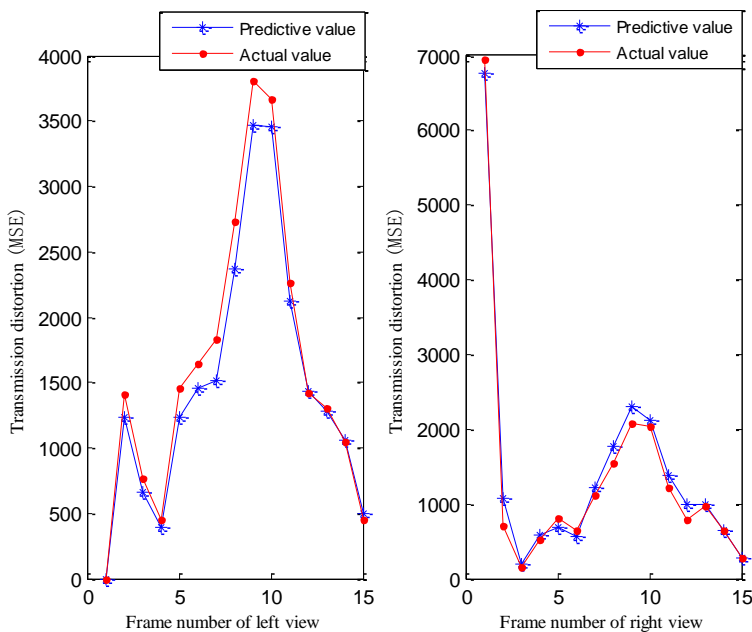
Figure 2. The total distortion of entire sequence due to loss of one book frame
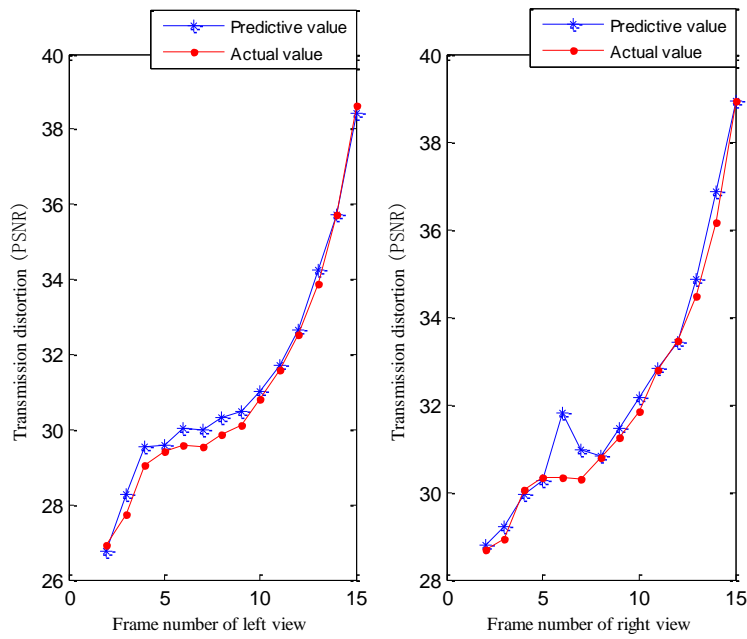
Figure 3.   Average PSNR of crowd after one frame lost
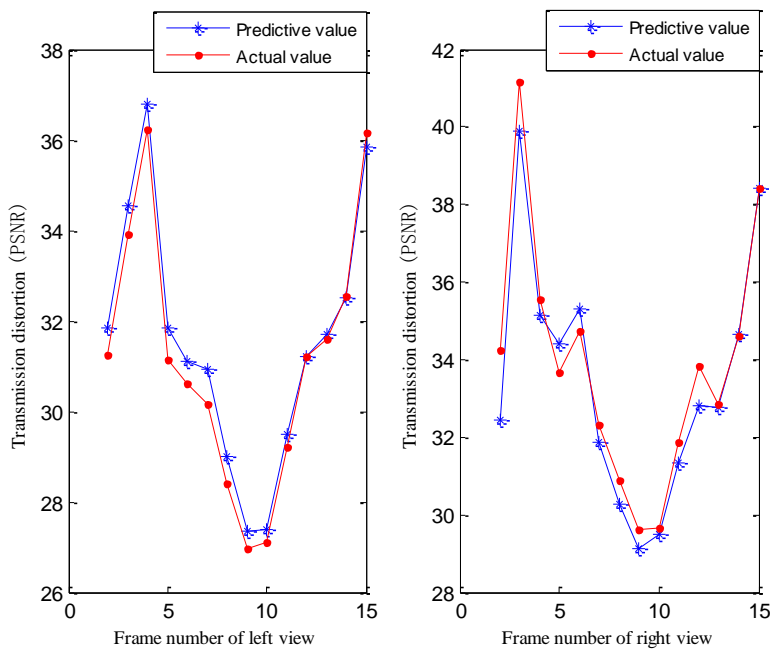
Figure 4.   Average PSNR of book after one frame lost

TABLE I.   PREDICTION ERROR OF THE FRAME IMPORTANCE FOR STEREO VIDEO

Sequence / View point	Book	Crowd	Flamenco2	Rollerblade	Aquarium
Predictive value of left view(MSE)	1478.21	1562.81	1188.03	1366.71	1552.55
Actual value of left view(MSE)	1616.67	1660.15	1165.66	1350.22	1545.68
Predictive value of right view(MSE)	1436.99	5520.92	1845.88	1543.22	1788.65
Actual value of right view(MSE)	1363.14	5694.97	1902.21	1550.36	1760.22
Predictive value of left view(PSNR)	29.43	29.24	30.34	29.35	29.65
Actual value of left view(PSNR)	29.10	29.03	30.59	29.44	29.77
Predictive value of right view(PSNR)	32.83	31.13	33.77	33.42	31.15
Actual value of right view(PSNR)	33.18	30.86	33.67	33.29	31.22

In order to better reflect the model accuracy, comparative results between predictive value and actual value was given in Table I.

Because the time correlation is stronger in a slow motion sequence, and frame loss for the frame and the subsequent frame is smaller, therefore the true value is

TABLE II.   THE AVERAGE PREDICTION ERROR OF THE FRAME IMPORTANCE FOR STEREO VIDEO

Sequence / View point	Book	Crowd	Flamenco2	Rollerblade	Aquarium	Average
MSE	5.22	4.78	5.76	6.95	4.18	5.38
MSE	8.81	3.26	3.73	3.25	7.53	5.24
PSNR/dB	0.94	0.65	0.95	0.78	0.70	0.80
PSNR/dB	1.08	0.97	0.94	0.72	0.77	0.90

greater than the slow motion sequence book. For the two sequences, the distortion value of first frame in right view is the largest, and the distortion size is related with the parallax view. This is because the time correlation of video sequence is greater than the spatial relativity, and it also related with the specific error concealment method. And the average frame importance in left view is larger than that in right view, this is because the right view frame loss will only affect its subsequent one while left view frame loss will affect all subsequent frames. From the figures, we can also know that the relationship between importance of frame and coding position is not linear which further demonstrates the conclusion that video contents have influences on the importance of frames.

In order to verify the accuracy of the model in this paper, we define $E(s,t)$ as the average prediction error of the frame importance as follows

$$E(s,t) = \frac{\sum_{t=0}^{N} \left| D_{sim}(s,t) - D_{act}(s,t) \right|}{\sum_{t=0}^{N} D_{act}(s,t)} \quad (s=0,1) \quad (12)$$

$D_{sim}(s,t)$ means distortion value, and $D_{act}(s,t)$ means actual experimental value. In order to verify the applicability of the model, we make experiments based on motion's sequence of different intensity and different parallax.

We examine five stereo video sequence, i.e., high motion sequences "crowd" and "rollerblade" whose camera distance are 20cm and 19.5cm; low motion sequence "book" whose camera distance is 6.5cm; rotating motion sequence "flamenco2", camera distance is 20cm; small-range motion sequence "aquarium" whose camera distance is 20cm.

In order to reduce the random error, experiments are made in the same condition. The average prediction errors of the frame importance for stereo video sequences are listed in Table I. We can see from table I that when MSE represents frame importance, average distortion error in left view is 5.97% and 4.97% in right view; when PSNR represents frame importance, average distortion error in left view is 0.80% and 0.90% in right view. Simulation results demonstrate that this model be applied to stereo video with different motion and disparity.

IV.   CONCLUSION

Based on the characteristics of stereo video sequences, the encoding and decoding algorithm and error concealment techniques, this paper proposes a perception oriented video frame importance differentiation model.

The experimental results show that the proposed model can be applied to stereoscopic video sequences with different motion intensity and disparity which can estimate the importance of each frame for terminal perception at code ends. Frame importance differentiation model can determine the stereo video encoding parameters (such as quantization parameter), optimize performance of the distortion rate, improve error control measures and transmission congestion control mechanism and guarantee for further optimization of stereoscopic video transmission system.

REFERENCES

[1] Zilly F, Kluger J, Berlin, et al. Production rules for stereo acquisition. *Proceedings of the IEEE,* 2011, 99(4) pp. 590-606.

[2] Urey H, Erden E, Surman P. State of the art in stereoscopic and autostereoscopic displays. *Proceedings of the IEEE,* 2011, 99(4) pp. 540-555.

[3] Feng Nina, Chang Yilin, Li Peng, et al. GOP-level video transmission method using data partitioning in H.264/AVC. *Journal of XiDian University (Natural Science),* 2011, 38(3) pp. 29-36.

[4] Gurler, C G, Koc Univ E, Istanbul. Flexible transport of 3-D video over networks. *Proceedings of the IEEE,* 2011, 99(4) pp. 694-707.

[5] Lin C H, Wang Y C, Shieh C K, et al. An unequal error protection mechanism for video streaming over IEEE 802.11 e WLANs. *Computer Networks,* 2012.

[6] Hartanto F, Sirisena H R. Hybrid error control mechanism for video transmission in the wireless IP networks. *10th IEEE Workshop on Local and Metropolitan Area Networks,* 1999 pp. 126-132.

[7] Lin C H, Shieh C K, Ke C H, et al. An adaptive cross-layer mapping algorithm for MPEG-4 video transmission over IEEE 802.11 e WLAN. *Telecommunication Systems,* 2009, 42(3-4) pp. 223-234.

[8] Ferré P, Agrafiotis D, Bull D. A video error resilience redundant slices algorithm and its performance relative to

other fixed redundancy schemes. *Signal Processing: Image Communication,* 2010, 25(3) pp. 163-178.

[9] Yang X, Zhu C, Li Z G, et al. An unequal packet loss resilience scheme for video over the Internet. *IEEE Transactions on Multimedia,* 2005, 7(4) pp. 753-765.

[10] Ohm J R, Sullivan G J. High Efficiency Video Coding: The Next Frontier in Video Compression [Standards in a Nutshell]. *IEEE Signal Processing Magazine,* 2013, 30(1) pp. 152-158.

[11] Djamah M, O'Shaughnessy D. Fine granularity scalable speech coding using embedded tree-structured vector quantization. *Speech Communication,* 2012, 54(1) pp. 23-39.

[12] Liu Y. Research of Progressive Fine Granular Scalable video coding technology *IEEE 3rd International Conference on Computer Research and Development (ICCRD),* 2011 pp. 155-15.

Xiaodong Wang received his B.S. degree in applied physical from Ningbo University, Ningbo, China, in 1992, and his M.S. degree in Zhejiang University, Hangzhou, China, in 2007. From 1992, he joined in Ningbo University, China. Since 2008, he works as a Associate Professor in Faculty of Information Science and Engineering. His research interests include network communications, 3D video processing and transmission system, etc.

Tengfei Wang is currently a college student and working towards his M.S. degree at Ningbo University, China. His current research interest includes stereoscopic video coding and network transmission.

Binbin Hu is currently a college student and working towards his M.S. degree at Ningbo University, China. His current research interest includes stereoscopic video error concealment and network transmission.

Gangyi Jiang received his M.S. degree from Hangzhou University in 1992, and received his Ph.D. degree from Ajou University, Korea, in 2000. He is now a professor in Faculty of Information Science and Engineering, Ningbo University, China. He is a member of the IEEE. His research interests mainly include digital video compression and communications, multi-view video coding, etc.

Lianjun Zhang received his B.S. degree in Computer science and technology from Ningbo University, Ningbo, China, in 2004, and his M.S. degree in Technology of Computer Application from Ningbo University, Ningbo, China, in 2012. His research interests include network communications, error concealment.

Research on Female Neck Rhizosphere Line of Virtual Mannequin Based on Point-Cloud Data

Pinying Gu [1, 2], Bingfei Gu [1, 2], and Guolian Liu [1, 2]

1. National Engineering Laboratory for Modern Silk, Ganjiang Road, Suzhou, Jiang Su, China
2. College of Textile and Clothing Engineering, Soochow University, Ganjiang Road, Suzhou, Jiang Su, China
Email: gpyappleyc@gmail.com, jstxgbf@aliyun.com, liuguolian@suda.edu.cn

Abstract—**With the rapid development of computer network and information technology, traditional clothing industry has taken a giant stride forward to computer information and digitization. In this paper, 250 3D point-cloud photos of young females were selected as subjects, and related characteristic points of neck rhizosphere (including front neck point, side neck point and back neck point) were determined. Then the height size, width size, thickness size and girth size of characteristic points were measured by the software named Imageware 12.1. At last, with software called Excel and SPSS, the height rules of characteristic points were analyzed, front and back neck rhizosphere were obtained by width and thickness sizes of characteristic points. The research in this paper has laid the foundation for building female neck rhizosphere line of virtual mannequin which provides datum line and basic sizes for 3D collar patterns.**

Index Terms—**Neck Rhizosphere Line; Point-Cloud Data; Virtual Mannequin; 3D Pattern**

I.   INTRODUCTION

With the development of computer information technology and social economic, people's demands for styles, qualities, fit and personalities are growing. But the 2D garment CAD technology can't meet the needs of garment industries, so many scholars conduct academic research in 3D garment CAD technology [1]. Now, the developing tendency of 3D garment CAD system mainly shows as follows: Virtual sample manufacture (or tailor made), remote clothing fitting (or virtual fitting, virtual shopping), merchandise planning, production mechanism building of quick response, and so on [2]. Mannequin is a reflection of a certain size and shape of the human body, which is an important tool for the production and presentation of clothing. 3D virtual mannequin is an important foundation for 3D garment CAD and apparel e-commerce.and it can not only exhibit different characteristics of the human body on the computer screen, but also ensure the virtual clothes to get the real wearing effect through the computer network. Hence, construction of 3D virtual mannequin is the key and focus to 3D garment CAD [3, 4, 5].

In connection with body and clothes, costume structural design needs to apply to human movement. Collar is the most attractive part of garment, and also an important part in garment construction. The collar shape has a great effect on overall garment, which not only reflects consistency with garment style, but also an important factor to show garment style [6, 7, 8]. Neck rhizosphere line is the foundation of the collar design. It reflects the neck shape related to collar design and decides the basic size for collar patterns. Hence, neck rhizosphere line of 3D virtual mannequin is very important for 3D collar patterns. However, the current research on 3D virtual mannequin cannot intuitively obtain neck rhizosphere line, we had to confirm neck rhizosphere line with other methods.

The position and shape of neck rhizosphere line are decided by three points named front neck point, side neck point and back neck point. In this paper, the reverse engineering software is used to conduct secondary development of the point-cloud data, and the body feature of female upper body was studied to establish the rules of automatically generating female neck rhizosphere line, which can be used to carry through the modeling of individual female upper body. Then, young female students of Soochow University were chose as research subjects and 3D body measurement, photo measurement and manual measurement methods were used to measure human body data. The details were as follows: with female point-cloud data measured and analyzed, the position rules of neck characteristic points (including front neck point, side neck point and back neck point) were found. Then, the neck rhizosphere line of 3D virtual mannequin was determined, which provided datum line and basic sizes for 3D collar patterns.

II.   CHARACTERISTIC POINTS OF NECK RHIZOSPHERE LINE

Neck rhizosphere means a round girth through front neck point, side neck point and back neck point, shown as Fig. 1. The neck rhizosphere line is the datum line for measuring neck rhizosphere, and that is an important basis to separate bodice and collar [9, 10]. To determine the shape and position of neck rhizosphere line, the position of front neck point, side neck point and back neck point should be found firstly.

A. The Front Neck Point

Front neck point is the endpoint on the first thoracic. On the front view of human body, a thumb-sized pit can be seen in the sternoclavicular joints, and it's just front

neck point. In this paper, the front neck point is marked as FNP.

B. The Side Neck Point

Side neck point is the cut-off point for the neck and shoulder, which is located in the root of the side neck. On the side view of human body, side neck point is a little off center point of neck root width. In this paper, side neck point is marked as SNP.

C. The Back Neck Point

Back neck point is the seventh cervial point, the most protruding point of back neck. On the back view of human body, back neck point can be easily seen. And it's mark as BNP in this paper [11, 12].

D. Characteristic Points Determined

With imageware, body point-cloud data were read, and with polygonize cloud operated, the body simulating mannequin can be obtained. Then each point can be found according to its characteristic, marked as FNP, SNP and BNP. Fig. 2 showed the determining process.
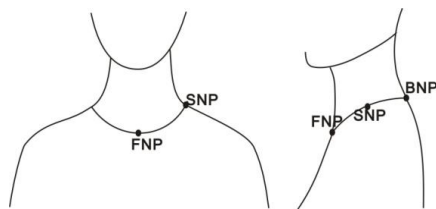

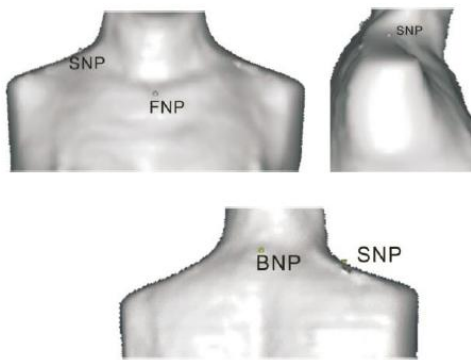Figure 1.   Schematic diagram of neck rhizosphere

Figure 2.   Locationsof characteristic points

III.   POINT-CLOUD DATA MEASUREMENT

In the previous research [13, 14, 15] about virtual female mannequin, the height rules and circumference rules of female characteristic parts have been obtained. The characteristic parts mainly include neck, chest, abdomen, waist and hip, and the position of side neck point has been found in neck study. Thus this paper focuses on the positions of front neck point and back neck point, that means researching FNP height and BNP height rules with the known height rule, and further researching circumference rules of front neck rhizosphere and back neck rhizosphere.

A. Subjects and Time

With the measuring time as 2014.2-2014.4, 204 female students from Soochow University were selected as research subjects, and they are aged from 18 to 26, height from 145cm to 178cm, weight from 40 Kg to 71 Kg.

B. Measuring Items

In this paper, with the software of imageware, the point-cloud data generated by 3D body scanner were readed and optimized. Then the related size of female neck rhizosphere line were measured, including height size (FNP height, BNP height and chest height), width size (SNP width), thickness size(FNP front thickness and BNP back thickness) and circumference size(front neck rhizosphere and back neck rhizosphere). To easily understand, these measuring items were explained as follows, and shown as Fig. 3.

(1) FNP height: In the front view of point-cloud picture, vertical distance from FNP to ground.

(2) BNP height: In the front view of point-cloud picture, vertical distance from BNP to ground.

(3) Chest height: In the front view of point-cloud picture, vertical distance from chest to ground.

(4) SNP width: In the left view of point-cloud picture, horizontal distance between 2 side neck points.

(5) FNP front thickness: In the left view of point-cloud picture, horizontal distance from FNP to vertical section of SNP width.

(6) BNP back thickness: In the left view of point-cloud picture, horizontal distance from BNP to vertical section of SNP width.

(7) Front neck rhizosphere: Along the neck rhizosphere direction, smoothly connected arcs from FNP to SNP.

(8) Back neck rhizosphere: Along the neck rhizosphere direction, smoothly connected arcs from SNP to BNP.
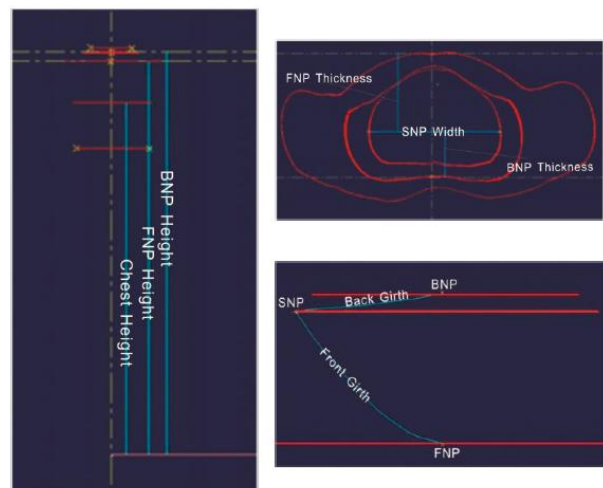
Figure 3.   Positions of measuring items

C. Measuring Method

The detailed measuring method is as follows: (1) After 3D point-cloud picture readed and optimized, find accurate positions of FNP, SNP and BNP. (2) In the front view of point-cloud picture, cut out the horizotal sections of FNP, BNP and BP respectively, then measure corresponding FNP height, BNP height and Chest height. (3) In the left view of point-cloud picture, measure the FNP front thickness, BNP back thickness and SNP width.

(4)Draw and measure front neck rhizosphere and back neck rhizosphere.

IV. DATA ANALYSIS

Data analysis includes two parts. One is height analysis, which mainly researches FNP height rule and BNP height rule by known chest height rule (Selecting chest height as a datum is due to that the chest is obvious feature of female parts, which can be found explicitly in 3D point-cloud pictures.). The other one is circumference simulation, which mainly simulates 3D sizes like front /back neck rhizosphere with 2D sizes including FNP front thickness, BNP back thickness and SNP width.

A. Height Data Analysis

With software of SPSS16.0, correlations of the FNP height, BNP height and chest height were analyzed and the scatter plots were drawn. Then regression equations about FNP height, BNP height and chest height were obtained by linear simulation.

1) Correlation Analysis About Characteristic Height of Neck Rhizosphere

In order to determine the trends and relevance that the dependent variable changes with the independent variable, it was necessary to verify through the scatter plot and correlation analysis. Table I showed the correlation analysis about FNP height, BNP height and chest height.

TABLE I.   THE CORRELATION ANALYSIS ABOUT FNP HEIGHT, BNP HEIGHT AND CHEST HEIGHT

	FNP height		BNP height	
	correlation coefficient	Sig.(2-tailed)	correlation coefficient	Sig.(2-tailed)
Chest height	0.952**	0.000	0.931**	0.000

** Correlation is significant at 0.01 level

Figure 4.   Scatter of FNP height and chest height

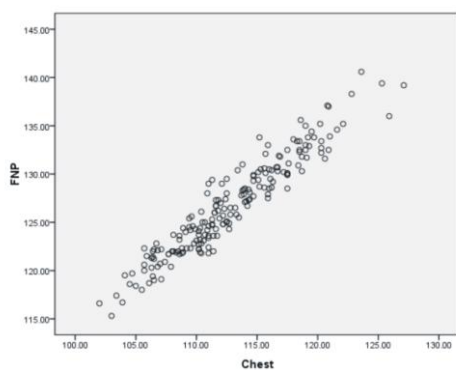Table I showed that: the correlation coefficient between FNP height and chest height is 0.952, and two-tailed level of significance is 0.000<0.01, indicating that the correlation between FNP height and chest height is very significant. And the correlation coefficient between BNP height and chest height is 0.931, and two-tailed level of significance is 0.000<0.01, so the correlation between BNP height and chest height is also very significant.

2) Scatter Plots About Characteristic Height of Neck Rhizosphere

Respectively putting "FNP height" and "BNP height" as dependent variable, "chest height" as independent variable, and the scatter plots were drawn as Fig. 4 and Fig. 5. Seen as scatter plots, with the increase of chest height, FNP height and BNP height showed a linear increase.
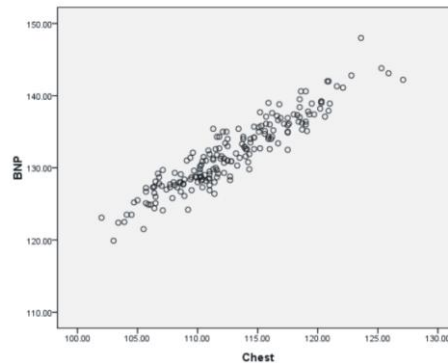
Figure 5.   Scatter of BNP height and chest height

Shown as the above analysis, linear regression method can be used to construct the relationships of FNP height, BNP height and chest height.

3) Regression Analysis About Characteristic Height of Neck Rhizosphere

The linear regression process of FNP height and chest height was shown in Table II, and the linear regression process of BNP height and chest height was shown in Table III.

Assuming the regression equation is $Y_1=a_1+b_1X_1$, then shown as Table II, a_1 is 16.732, b_1 is 0.975, thus regression equation is $Y_1=16.732+0.975X_1$. Y_1 indicates FNP height, X_1 indicates chest height. And the regression constant is 16.732, two-tailed level of significance is 0.000<0.01, significant symbol is "***", so the regression constant is very significant. Regression coefficient is 0.975, two-tailed level of significance is 0.000<0.01, significant symbol is "***", so the regression coefficient is very significant.

TABLE II.   REGRESSION ANALYSIS OF FNP HEIGHT AND CHEST HEIGHT

	Un-standardized coefficients		Standardized coefficient	T	Sig.
	B	Std.Error	Beta		
Constant	16.732	2.556		6.545	.000
Chest height	.975	.023	.952	43.056	.000

a. Dependent variable: FNP

Assuming the regression equation is $Y_2=a_2+b_2X_1$, then shown as Table III, a_2 is 25.720, b_2 is 0.942, thus regression equation is $Y_2=25.72+0.942X_1$. Y_2 indicates BNP height, X_1 indicates chest height. And the regression constant is 25.72, two-tailed level of significance is 0.000<0.01, significant symbol is "***", so the regression constant is very significant. Regression coefficient is 0.942, two-tailed level of significance is 0.000<0.01, significant symbol is "***", so the regression coefficient is very significant.

B. Fitting Analysis of Circumference

1) Correlation Analysis of Width, Thickness and Circumference

In previous research, fitting circumference with corresponding width and thickness of characteristic part is feasible. Thus in this paper, FNP front thickness and SNP width were used to simulate front rhizosphere, and BNP back thickness and SNP width were used to simulate back rhizosphere. Table IV showed the correlation of FNP front thickness, SNP width and front neck rhizosphere, Table V showed the correlation of BNP back thickness, SNP width and back neck rhizosphere.

TABLE III. REGRESSION ANALYSIS OF BNP HEIGHT AND CHEST HEIGHT

	Un-standardized coefficients		Standardized coefficient	T	Sig.
	B	Std.Error	Beta		
Constant	25.720	3.007		8.553	.000
Chest height	.942	.027	.931	35.367	.000

a. Dependent variable: BNP

TABLE IV. THE CORRELATION OF FNP FRONT THICKNESS, SNP WIDTH AND FRONT RHIZOSPHERE

	FNP front thickness		SNP width	
	correlation coefficient	Sig.	correlation coefficient	Sig.
Front neck rhizosphere	0.744**	0.000	0.583**	0.001

** Correlation is significant at 0.01 level

As Table IV shown: the correlation coefficient of front neck rhizosphere and FNP front thickness is 0.744, and two-tailed level of significance is 0.000<0.01, indicating their correlation is very significant. In addition, the correlation coefficient of front neck rhizosphere and SNP width is 0.583, and two-tailed level of significance is 0.000<0.01, which means front neck rhizosphere and SNP width have a significant correlation. Thus, fitting front neck rhizosphere with FNP front thickness and SNP width is possible.

TABLE V. THE CORRELATION OF BNP BACK THICKNESS, SNP WIDTH AND BACK NECK RHIZOSPHERE

	BNP back thickness		SNP width	
	correlation coefficient	Sig.	correlation coefficient	Sig.
Back neck rhizosphere	0.683**	0.000	0.547**	0.000

** Correlation is significant at 0.01 level

Table V showed that: the correlation coefficient of back neck rhizosphere and BNP back thickness is 0.683, and two-tailed level of significance is 0.000<0.01, indicating their correlation is significant. In addition, the correlation coefficient of back neck rhizosphere and SNP width is 0.547, and two-tailed level of significance is 0.000<0.01, which means front neck rhizosphere and SNP width have a significant correlation. Thus, fitting back neck rhizosphere with BNP back thickness and SNP width is possible.

2) Regression Analysis of Width, Thickness and Circumference

The regression of FNP front thickness, SNP width and front neck rhizosphere was shown in Table VI. Assuming the regression equation is $Y3=a+bX2+cX3$, then a is 7.218, b is 0.570 and c is 0.120, thus regression equation is $Y3=7.218+0.57X2+0.12X3$. Y3 indicates front neck rhizosphere, X2 indicates FNP front thickness, X3 indicates SNP width. And the regression constant is 7.218, two-tailed level of significance is 0.000<0.01, significant symbol is "***", so the regression constant is very significant. Regression coefficient of FNP front thickness is 0.570, two-tailed level of significance is 0.000<0.01, significant symbol is "***", indicating the regression coefficient is very significant. Regression coefficient of SNP width is 0.120, two-tailed level of significance is 0.039<0.05, significant symbol is "**", indicating the regression coefficient is significant.

TABLE VI. THE REGRESSION ANALYSIS OF FNP FRONT THICKNESS, SNP WIDTH AND FRONT NECK RHIZOSPHERE

	Un-standardized coefficients		Standardized coefficient	T	Sig.
	B	Std.Error	Beta		
Constant	7.218	.871		8.290	.000
FNP front thickness	.570	.068	.516	8.367	.000
SNP width	.120	.058	.128	2.080	.039

a. Dependent variable: Front neck rhizosphere

The regression of BNP back thickness, SNP width and back neck rhizosphere was shown in Table VII. Assuming the regression equation is $Y4=a+bX4+cX3$, then a is -0.437, b is 0.680 and c is 0.522, thus regression equation is $Y4=-0.437+0.68X4+0.522X3$. Y4 indicates back neck rhizosphere, X4 indicates BNP back thickness, X3 indicates SNP width. And the regression constant is -0.437, two-tailed level of significance is 0.049<0.05, significant symbol is "**", so the regression constant is significant. Regression coefficient of BNP back thickness is 0.680, two-tailed level of significance is 0.000<0.01, significant symbol is "***", indicating the regression coefficient is very significant. Regression coefficient of SNP width is 0.522, two-tailed level of significance is 0.000<0.01, significant symbol is "***", indicating the regression coefficient is very significant.

TABLE VII. THE REGRESSION ANALYSIS OF BNP BACK THICKNESS, SNP WIDTH AND BACK NECK RHIZOSPHERE
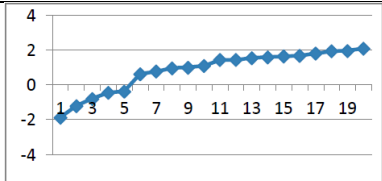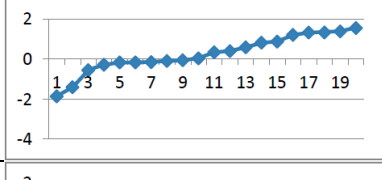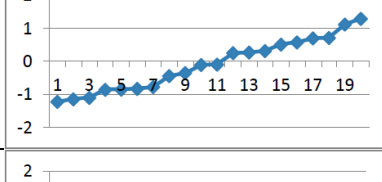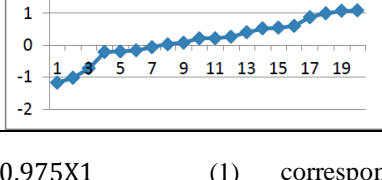
	Un-standardized coefficients		Standardized coefficient	T	Sig.
	B	Std.Error	Beta		
Constant	-.437	.923		-.474	.049
BNP back thickness	.680	.072	.509	9.513	.000
SNP width	.522	.058	.477	8.924	.000

a. Dependent variable: Back neck rhizosphere

C. Summary

By data analysis above, equations as follows can be obtained:

TABLE VIII.   THE ERROR ANALYSIS

	Error distribution	Upper limit error	Lower limit error
FNP height		-1.89	2.07
BNP height		-1.85	1.56
Front neck rhizosphere		-1.22	1.28
Back neck rhizosphere		-1.18	1.09

$$Y1 = 16.732 + 0.975X1 \qquad (1)$$

$$Y2 = 25.72 + 0.942X1 \qquad (2)$$

$$Y3 = 7.218 + 0.57X2 + 0.12X3 \qquad (3)$$

$$Y4 = -0.437 + 0.68X4 + 0.522X3 \qquad (4)$$

Y1 represents FNP height, Y2 represents BNP height, Y3 represents front neck rhizosphere, Y4 represents back neck rhizosphere, X1 represents chest height, X2 represents FNP front thickness X3 represents SNP width, X4 represents BNP back thickness.

V.   DATA VALIDATION

In order to verify if the regression equations are accurate, 20 other samples were selected for a validation measure. With software of imageware, data of chest height, FNP front thickness, BNP back thickness and SNP width were measured. Then these data were brought into equations above, and corresponding FNP height, BNP height, front neck rhizosphere and back neck rhizosphere were calculated. The error analysis of measured and calculated values was shown in Table VIII. Table VIII showed that: the fitting errors of FNP height and BNP height are within ± 2cm, the fitting errors of front neck rhizosphere and back neck rhizosphere are within ± 1.3cm, which meet the needs of apparel structure design. Thus, the equations obtained in this paper are effective.

VI.   CONCLUSION

In this paper, with the measurement and research for female point-cloud data, the rules between FNP/BNP height and chest height were found, and the corresponding equations were obtained. Furthermore, the front and back neck rhizosphere were simulated by FNP front thickness, BNP back thickness and SNP width, which laid the foundations for female neck rhizosphere line of 3D virtual mannequin and provided datum line and basic sizes for 3D color patterns.

REFERENCE

[1] BF Gu. "Research on 3D Virtual Modeling of Female Upper Body Based on Point-Cloud Data," *Master Dissertation. Soochow University:* 2012.
[2] BF Gu, and GL Liu. "Automatic Modelling of the Lower Bodies of Young Females Based on Digital Photographs," *Journal of Fiber Bioengineering and Informatics,* 7: 2 (2014) pp. 247-260.
[3] CY Kuang, KC Ji, R Jin and GL Liu. "Prediction of the Main Girths of Young Female Based on 3D Point Cloud Data," *Journal of Fiber Bioengineering and Informatics:* 2012; 3 pp. 289-298.
[4] P Xia, FP Ou, C Yang and J Yao. "The Modeling of the Individual Mannequin," *Journal of Sichuan University (Engineering Science Edition).* 2009, 6 pp. 227-232.
[5] J Chen. "Research and Realization for Collar CAD Technology," *Master Dissertation. Zhe Jiang University:* 2006.

[6] LJ Jiang. "Research on Collar 3D Construction Digital Technology Based on Neck Characteristics," *Doctoral Dissertation. Dong Hua University:* 2010.

[7] JJ Zhao. "Study on Impact Factors of Stand-up Collar's Upper Arc Length Based on the Three-dimensional Virtual Technology," *Master Dissertation. Dong Hua University:* 2013.

[8] HX Qi. "Research on Evaluation Model of Clothes Fit Based on Individual Virtual Mannequin," *Doctoral Dissertation. Dong Hua University:* 2011.

[9] PY Gu, T Chen, L Huang, HY Kong and GL Liu. "Study on Calculating Methods of Young Women Body Circumferences Based on 3D Point-cloud Data," *Advanced Materials Research.* 2011 pp. 2010-2013.

[10] LH Luo and Shi XP. "Study on Automatic Acquisition of Neck Circumference Line Based on E-MTM," *Computer Applications and Software.* 2010, 7 pp. 77-79.

[11] CY Kuang. "Research on Non-contact Measurement of Young Men's Bodies Based on Digital Image," *Master Dissertation. Soochow University:* 2008.

[12] HY Kong. "Research on 3D Virtual Modeling of female lower body and prototype patterns of tight pants," *Master Dissertation. Soochow University:* 2012.

[13] BF Gu and GL Liu. "Research on Auto human body measurement and Application System," *Journal of Applied Sciences.* 2013, 13(18) pp. 3691-3697.

[14] BF Gu and GL Liu. "Research on Structural Design of Pants and pattern generation system," *Journal of Multimedia.* 2013, 8, 6 pp. 809-815.

[15] PY Gu, Q Qian, T, Chen L Huang, HY Kong and GL Liu. "Study on Pattern Design Method of Women Tight Skirts Based on 3D Point-cloud Data," *Journal of Fiber Bioengineering and Informatics.* 2012, 1 pp. 85-93.

Pinying Gu Suzhou city, China, was born in May 1990. Be in Master Course of garment design and engineering of Soochow University from 2012. The major field focuses on the research of digitized human engineering and automatic generation of garment pattern.

Paper Similarity Detection Method Based on Distance Matrix Model with Row-Column Order Penalty Factor

Li Jun [1, 2]
1. Hubei University of Technology, Wuhan, China
2. HuaZhong University of Science and Technology, Wuhan, China
Email: 50331615 @qq.com

Han Yaqing
Hubei University of Technology, Wuhan, China
Email: 7137100 @qq.com

Pan Junshan
HuaZhong University of Science and Technology, Wuhan, China
Email: 58510148@qq.com

Abstract—**Paper similarity detection depends on grammatical and semantic analysis, word segmentation, similarity detection, document summarization and other technologies, involving multiple disciplines. However, there are some problems in the existing main detection models, such as incomplete segmentation preprocessing specification, impact of the semantic orders on detection, near-synonym evaluation, difficulties in paper backtrack and etc. Therefore, this paper presents a two-step segmentation model of special identifier and Sharpley value specific to above problems, which can improve segmentation accuracy. In the aspect of similarity comparison, a distance matrix model with row-column order penalty factor is proposed, which recognizes new words through search engine exponent. This model integrates the characteristics of vector detection, hamming distance and the longest common substring and carries out detection specific to near-synonyms, word deletion and changes in word order by redefining distance matrix and adding ordinal measures, making sentence similarity detection in terms of semantics and backbone word segmentation more effective. Compared with the traditional paper similarity retrieval, the present method has advantages in accuracy of word segmentation, low computation, reliability and high efficiency, which is of great academic significance in word segmentation, similarity detection and document summarization.**

Index Terms—**Paper Detection; Segmentation; Similarity Comparison; Distance Matrix; Model Modular-2 Arithmetic**

I. INTRODUCTION

With the development of network and information technology, information sharing and dissemination is becoming more and more convenient, since it has provided an ideal platform for academic researchers. However, it also bred a series of academic misconducts, such as plagiarism, practicing fraud and so on. Besides complete plagiarism, cases like plagiarizing others' achievements by transposition, paraphrasing and synonym replacement are also very common. These behaviors cause damage to academic quality. Consequently, aiming at eliminating these academic misconducts and improving the quality of paper, paper similarity detection which depends on word segmentation, similarity detection, document summarization and other means are quite significant.

Since word segmentation is the basis and premise of Chinese text similarity calculation, and the adoption of segmentation algorithm of high efficiency can greatly improve the accuracy of text similarity calculation, it is essential to keep on exploring new segmentation algorithm on the basis of existing ones, and improve the integrity and accuracy of word segmentation so as to make the comparison of similarity between texts more accurate, and thus to provide decision support for related business.

Since Chinese has a large volume and variable semantics, copy detection of Chinese paper is more complex and difficult than that of English paper, and the basic resource (such as corpus) for Chinese language processing such as text detection is relatively insufficient, which makes it impossible to apply mature technology and research achievements abroad directly. As a result, information processing of Chinese by computers is more difficult than that of western languages. Meanwhile, there are significant differences between Chinese and English segmentation because Chinese text is a continuous string of large character set, which means no specific separation mark exists among Chinese words, while English text is a fully separated string of small character set by space. Therefore, similarity comparison between Chinese texts

must be inspected based on segmentation systems, and the accuracy of word segmentation determines that of paper similarity calculation. Despite the achievements of word segmentation algorithm, there still exist problems as follows:

(1) Disunity in segmentation specification and inconsistency in dictionary design;

(2) Incompleteness of segmentation algorithm;

(3) Insufficient ambiguity correction mechanism.

Based on similarity calculation, paper similarity detection manages to calculate the similarity between texts automatically. The calculation of text similarity has been widely applied in such fields as retrieval, machine translation, question answering systems and text mining, which is a basic and key problem and has long been a hot topic and difficulty for researchers. At present, many scholars both at home and abroad are studying text similarity calculation problem and have put forward some solutions.

In 1993, Manber [1] from Arizona University proposed the concept of approximate exponential, which was the earliest detection technology and was adopted to measure string similarity among documents. Later, Bao Junpeng [2] and others proposed document copy detection method based on semantic sequence which emphasized position information of words and improved the detection accuracy. Afterwards they proposed corresponding detection model (SSK) [3] which is suitable for copy detection without replacement of words.

In 2005, Jin Bo [4] from Dalian University of Technology firstly extended the scope of text similarity calculation into paragraph based on HowNet semantic similarity. Then he extended paragraph similarity calculation into chapter and proposed text (including words, sentences and paragraphs) similarity calculation formula and algorithms. In addition, he proposed long text similarity copy detection algorithms [5] in 2007 which made it possible to calculate the coverage of similar semantic sequence set by semantic sequence similarity relation and choose each overlap values with the minimum entropy for plagiarism recognition and retrieval.

Hung Chenghui et al. [6, 7] from Zhong Shan University proposed a text similarity calculation method by combining word semantic information with TF-IDF where sentence was regarded as vector space composed of independent entry, and the matching problem of document information was transformed into that of vector in the its space and then the similarity was obtained by dot product method and cosine method. However, TF-IDF which is based on vector space model also has disadvantages. Firstly, it is a statistical-based method and only when the text includes reaches certain length can some words occur repeatedly, which will then reflect its statistical results. Secondly, this method only takes statistical lexical features in context without the semantic information, and thus produces some limitations. Just as Salton form Cornell University says, there is no strict theoretical basis for calculating vector similarity by angle cosine.

To enhance the performance of result merging for distributed information retrieval, a novel merging method was put forward by Wang Xiuhong [8] from Jiang Su University and others, which was based on relevance between retrieved results and query.

In 2009, Nie Guihua [9] from Wuhan University of Technology proposed systematic frame model of ontology-based thesis copy detecting system which described framework of paper copy detecting system from three layers: ontology access layer, ontology represent layer and ontology map layer. Semantic and ontology technology was utilized to discuss the build of paper ontology and calculation of paper similarity.

Zhang Huanjiong [10] from Beijing University Of Posts And Telecommunications constructed a new formula to calculate text similarity from Hamming calculation formula based on Hamming distance theory. It has the advantage of simplicity and rapidity but ignores the effects of unequal-length text and word order.

Later in 2011, Chen Yao-Tsung [11] from Tai Wan proposed a method adopting chi-square statistics to measure similarities of term vector for texts, which reduces the miss rate of similarity detection.

Also, Sánchez-Vega [12] from Mexico proposed a rewriting exponential model based on finite state machine which manages to detect common actions performed by plagiarists such as word deletion, insertion and transposition, but fails to deal with near-synonyms.

Through above analysis, we can still find some problems in main detection methods at present as follows:

(1) The statistical property of words in the context is taken into consideration while semantic information is ignored.

(2) The effects of unequal-length text and word order are not fully considered.

(3) There is limitation in recalling paper source by backtrack.

On the other hand, similarity calculation has different requirements for different applications, such as paper-level, paragraph-level, sentence-level, word-level and morpheme-level, and is often represented by formula or model of similarity calculation. However, with the evolution of plagiarism mechanism and diversification of expressions, problems such as large computation and difficulties in feature extraction and correct understanding of evaluation standards in paper-level and paragraph-level similarity detection still exists. As for word-level similarity, it may deviate greatly from actual value because of its small particle size. In addition, the space and time complexity of detection algorithm are quite large, making it difficult to obtain ideal effect from its application into huge amounts of paper.

A. Contribution

We summarize our contribution as below:

Starting with word segmentation and text similarity research, through integrating several main detection models at present, this paper builds the model specific to plagiarism, replacement of words and transposition based on the analysis of existing detection technologies.

Firstly, a two-step segmentation model is proposed, where we segment words by properties in step one and establish cooperative game model for roots in step two to calculate Sharpley Value, maximizing N-GRAM value.

Secondly, an evaluation score model of new words and popular words in search engine is proposed to make corpus of segmentation preprocessing more complete.

Finally, a distance matrix model with row-column order penalty factor is proposed. This model integrates the characteristics of vector detection, hamming distance and the longest common substring and carries out detection specific to near-synonyms, word deletion and changes in word order by redefining distance matrix and adding ordinal measures, making sentence similarity detection in terms of semantics and backbone word segmentation more effective and the returned results of retrieval better meet the requirements of users. Different from the traditional paper similarity retrieval, the present method adopts modular-2 arithmetic with low computation whose effectiveness and practicability have been confirmed by corresponding experiment.

II. DISTANCE MATRIX MODEL WITH ROW-COLUMN ORDER PENALTY FACTOR

A. Segmentation Based on Special Identifier

In the process of text similarity calculation, segmentation preprocessing must be carried out in the first place, since participial accuracy determines that of paper similarity calculation. There are several common algorithms for Chinese segmentation at present, such as unsupervised segmentation, mechanical word segmentation based on dictionary, segmentation based on linguistic model and segmentation based on character tagging. The first algorithm judges the correlation between two characters by computing their mutual information in corpus, which has good effect on high frequency words but is affected by threshold coefficient [13]; the second one does not have universality because it is affected by specialties of dictionary, and the third algorithm is the most commonly used method which is under research and development. Generally, Chinese Segmentation is confronted with 2 challenges: ambiguity and identification of new words.

According to the recently published Chinese Grammar Questions (Another version of Chinese Grammar) compiled by Professor Xing Fuyi, Chinese words are classified into three categories and eleven small classes, in which nouns, verbs, adjectives, numerals, quantifiers and pronouns are classified as notional words and adverbs, prepositions, conjunctions, auxiliary words, onomatopoeias and interjections are classified as functional words, which shows that POS features of text can be used as basis of recognition of special identifiers.

From discussion above, it can be concluded that special identifiers are words or symbols of specific significance and function, which are composed of non-Chinese characters such as punctuations, unit symbols, mathematical symbols, numerical symbols and letters, and Chinese characters such as special symbols. Here, a systematic analysis of the semantic and POS

features of text is conducted by reference to Chinese Grammar, and then a special identifier dictionary is constructed to recognize special identifiers in the text and confirm them.

B. The Segmentation Model Based on Sharpley Value

After the segmentation with special identifiers, its result probably needs further optimization. Thereupon, we further divide partial results into bi-roots, tri-roots and multi-roots. If root A, B, C, and D form the following 5 combination patterns: {AB\ CD}{A\ BC\ D}{ABC\D}{A\BCD}{ABCD}, the segmentation should be relevant to the prepositional and postpositional semanteme of the aforesaid combinations. Obviously, different combinations have different maximums of N-GRAM, and N-GRAM is widely applied in the field of Hanzi speech recognition. When we combine Hanzi into sentences, we could locate those sentences of maximum probability by calculation. Suppose some part of the sentence M needs further division, then S={a1, a2, …… an}, Here, ai refers to different segmenting combination (mono-root, bi-root and tri-root and etc.,) Pre_M refers to the pre-positional semanteme, and Fol_M the post-positional semanteme. Besides, every word segment within M maintains the relationship of cooperative game. The segmentation thereof makes the final N-GRAM a maximum. Consequently, Max{V(Pre_S)+ V(S)+ V (Fol_S)}. According to Shapley Theory, a unique value function exists:

$$\varphi_i(v) = \sum_{S \subseteq N\{i\} i \in S} \gamma_i(S)(v(S) - v(S - \{i\}))$$

$i \in N$, within which $\gamma_i(S) = \dfrac{(s-1)!(n-s)!}{n!}$

Accordingly, the assigned value gained by members of cooperative alliance becomes Shapley Value, denoted as $\varphi(v) = (\varphi_1(v), \varphi_2(v), \varphi_3(v)......, \varphi_n(v))$, within which, $\varphi_i(v)$ indicates the benefit gained by member i and is measured by N-GRAM. In different combination, the individual benefit differs from each other. Therefore, the ultimate segmenting result is derived from the N-GRAM value of the maximum sentence-level.

C. Evaluation Score Model of New Words and Popular Words in Search Engine

The cyber vocabulary expands at a high speed, and search engine has unique advantages in processing new words, popular words and common ambiguous segments because of its wide applications. Naturally, assessment of new words and popular words can be conducted with the help of search engines' results.

$$\text{Value(Word X)} = \frac{Number \text{ of pages on Baidu}}{100,000,000} \times k_{search}$$

Within which, N represents the number of pages on Baidu and the denominator 100,000,000 is the maximum number of pages after searching the commonly used words on Baidu.

D. Sentence-level Corpus and Digital Signature

Automatic summarization is divided into key phrase

extraction and text generation, which is used to explain and summarize the content of the original text with a brief description so as to accelerate people understanding information and solving the problem of information overloaded effectively. On the basis of statistics, paper-level and paragraph-level summary is composed of key sentences extracted from the text by Bayesian method and Hidden Markov Model according to word frequency and position information of the text, which is mainly suitable for technical documents with standardized format, and for other types of document, it needs further improvement in plagiarism-detection. Because of its unidirectional mapping mechanism, summary method is of good effect on complete paragraph plagiarism, which is rarely used in actual situation. In addition, it is of great complexity to extract key information with representativeness for paper-level and paragraph-level summary. However, the theory of summary can be used to create a digital signature which is composed of paper title and author information for each sentence and add it to sentence-level corpus for the subsequent backtrack.

A paper is composed of multiple sentences and after the segmentation is conducted, each sentence can be regarded as a set of words and then a sentence-level similarity comparison of the paper to be detected with the compared paper will be conducted. Firstly, the segmentation results of compared paper will be added into database, and secondly each sentence in the paper to be detected with that in the database compared. In order to look up the paper source by backtrack, a digital signature composed of paper title and author information is appended to each sentence. Thus in the similarity comparison process, if two sentences are judged to have a high similarity, their corresponding papers can be retrieved quickly.

E. General Similarity Comparison Model with Sequence Factor

VSM (such as IF-IDF) transforms matching problems of document information to that of vector in the vector space with a comprehensive consideration of term frequency (TF) in all texts and the term' ability to distinguish catalogs (IDF) [14, 15]. Hamming Distance adopts modular-2 arithmetic, avoids a mass of multiplication during similarity process in Euclid space, and obtains a high calculation speed. All these methods have their advantages, but at the same time, they have such problems as slow paper backtrack and ignorance of word order. For instance, in the following sentences: (a) "lexical collocation methods are various"; (b)" lexical types are various"; (c)" English lexical collocation methods are various", the similarities which are obtained by VSM, Hamming Distance and the longest common substring, respectively, are approximate if there is no consideration of word order, but obviously, the similarity of sentence a and sentence c is higher than that of sentence a and b. So comparison of sentence-level similarities with consideration of word order will obtain better effects.

Distance matrix model with row-column order penalty factor in this paper integrates the characteristics of VSM, hamming distance and the longest common substring, and realizes the transformation among these models through different parameter configurations.

In this model, paper to be detected, called A, and compared paper, called B, are divided into two sets of sentences, and each is a segmentation vector. So after the division, paper A is described as:

$LA = \{ A_1, A_2, A_3 \ldots A_n \}$,paper B is described as: $LB = \{ B_1, B_2, B_3 \ldots B_n \}$, and if the two vectors have a different number of elements, we can add some empty string elements into the one with fewer elements until it has as many elements as the other, and then conduct the following calculations:

$$LA' \otimes LB = \begin{pmatrix} A_1 \\ A_2 \\ \ldots \\ A_n \end{pmatrix} \otimes \begin{pmatrix} B_1 & B_2 & \ldots & B_n \end{pmatrix}$$

$$Sim(LA', LB') = \frac{|LA' \otimes LB|}{k\sqrt{\sum_{i=1}^{n} A_i^2 * \sum_{j=1}^{n} B_j^2}} * f(TA, TB)$$

$| LA' \otimes LB |$ is the maximum number of words that are identical in two sentences, which is similar to the longest common substring, and the difference is that the effects of multiple identical substrings are taken into consideration in this method. XOR is adopted in the calculation of $LA' \otimes LB$, whose result is an N-order sparse matrix. When the element from vector LA is multiplied by the element form vector LB, the result will be 1 if they are identical. Otherwise it will be 0. The first non-zero elements in rows are counted to create a row-sequence table, called TA, saving the row numbers of those elements that have identical elements in LB, and a column-sequence table ,called TB, saving the column numbers of those elements that have identical elements in LA. TA and TB represent the word orders of the longest substrings in LA and LB, respectively. Then the similarity of TA and TB is compared by utilizing Euclidean distance and the following function $f(TA, TB)$ is called:

$$f(TA, TB) = sqrt(\sum | TA_i - TB_i |^2)$$

In which, i is the serial number of TA and TB, n is the elements number of TA and TB, and the result of the function will be regarded as the row-column order penalty factor.

III. EXPERIMENTS

A. Segmentation Preprocessing

Experimentally, we select a news page randomly, and conduct individual segmentation by reference to our segmentation model. As shown in Figure 1 (two phases presented), phase 1 refers to the segmentation result according to special identifiers whereas phase 2 indicates the result of Sharpley value segmentation. After these two phases, the accuracy rate reaches 97.2%, and the recall rate 90.2%, satisfactory.
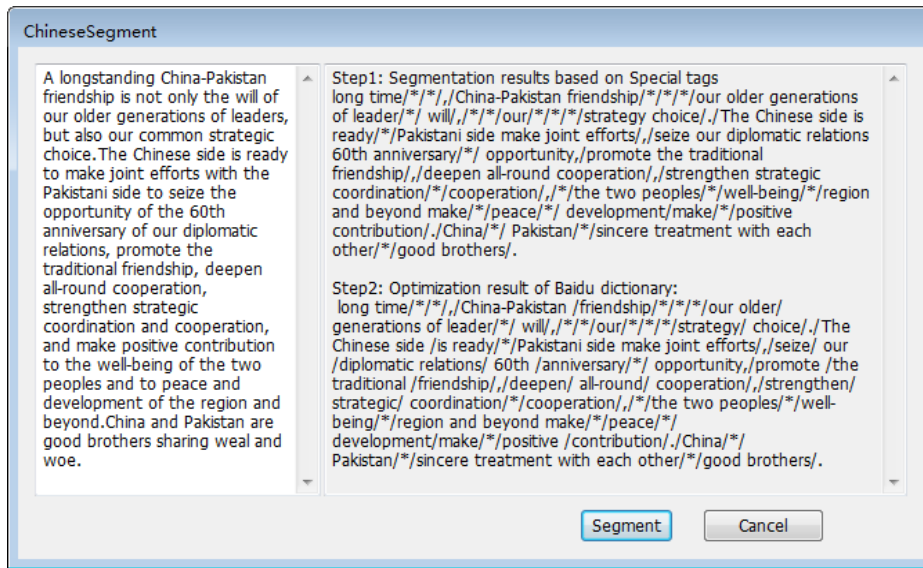
Figure 1.   Segmentation result based on special identifier and Baidu Dictionary

B. The Recognition of New Words and Hot Words

When we input "Gangnam style" and "Gannam style" into Baidu search engine for new words recognition, the former turns out to be: "appears in 17,000,000 relevant pages", and k_{search} is 18403, while the latter turns out to be 1,790, and k_{search} is 10. According to evaluation score, the former should be a hot word.

$$\text{Value(Gangnam style)} = \frac{17,000,000}{100,000,000} * 18403 = 3128.5$$

$$\text{Value(Gannam style)} = \frac{1790}{100,000,000} * 10 = 0.000179$$

Through repetitive analysis of the result, those words whose values exceed 1000 could be added into vocabulary as hot words.

C. Research on Sentence-Level Corpus and Digital Signature

This paper adopts MD5 algorithm in experiment, which can process messages of any length and produce a message summary with a fixed length, so the massage summary obtained through the input of paper title and author information is added into the sentence-level corpus as a digital signature. Figure 2. shows the sentence-level summary.

string	Objects Similarity Correlation Calculate Become Data Mining Information Extract Field Problem
title	Ontology Based Semantic Similarity and Ralatedness Measures Review
author	Liu Hong-zhe Author
hash value	A8DDED5E696AE135F5C9D6EDFB016ADE

Figure 2.   Result of sentence-level summary

D. General Similarity Comparison Model with Sequence Factor

The following words are selected to analyze similarities:

we entertained Beijing visitors passionately (segmentation results of compared sentence in corpus)
101 203 254 357 656
Beijing visitors passionately entertained we (Pa: sentence to be detected)
we passionately entertained Beijing visitors(Pb: sentence to be detected)
After segmentation, the sentence vectors are:
s= [101 * 203 * 254 * 357 * 656];
a= [254 * 357 * 656 * 203 * 101];
b= [101 * 656 * 203 * 254 * 357];
'*'means there may be other words among the given ones. Through the vector multiplication and deletion of rows of all zero members, a matrix is obtained as follows:

$$a' \otimes s = \begin{bmatrix} 254 \\ 357 \\ 656 \\ 203 \\ 101 \end{bmatrix} \otimes \begin{bmatrix} 101 & 203 & 254 & 357 & 656 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

From the result, TA of sentence $a' \otimes s$ can be expressed as: (3, 4, 5, 2, 1) and TA of sentence $b' \otimes s$ can be expressed as: (1, 5, 2, 3, 4). With the assumption that word order of sentence Ps is: (1, 2, 3, 4, 5), the word order curve graph of three sentences is shown in Figure 3. Figure 4. and Figure 5. show Euclidean distances of Pa and Pb, respectively.

After calculation, the results of f (Pa, Ps) and f (Pb, Ps) are: 5.6569 and 3.4641, indicating that penalty factor of Pb is smaller than that of Pa. Namely, similarity between Pb and Ps is higher than that between Pa and Ps, which obviously agrees with manually assessing results, under

the condition that the number of words that are the same between Pa and Ps equals to that between Pb and Ps. If $f(TA,TB)$ is 1, which means word order is ignored, the model evolves into VSM model with Dice coefficient and if the weight of each word is ignored and multiplication is set as the algorithm, this model evolves into hamming distance model, where k is a constant related to vector, if k is set to:

$$\frac{1}{\sqrt{\sum_{i=1}^{n} A_i^2 * \sum_{j=1}^{n} B_j^2}}$$ and the number of words that are exactly

the same in the compared sentence and the detected sentence is defined as the rank, this model evolves into the longest common substring under the condition that word order is ignored. Otherwise it is the longest common substring affected by order penalty factor.
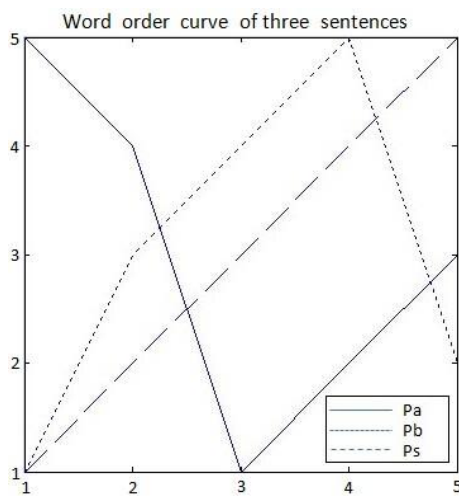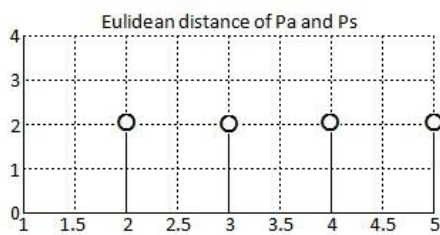

Figure 3.   Word order of three sentences

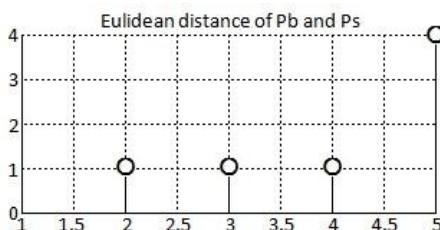Figure 4.   Euclidean distance of Pa and Ps

Figure 5.   Euclidean distance of Pb and Ps

E. Processing of Replacement of Words and Paraphrasing

For the phenomenon of common replacement of words and paraphrasing, near-synonym processing is adopted, and the corresponding dictionary of near-synonyms,

which are mainly from Chinese Synonyms Dictionary and those that are the same word in English, is build. In calculation of (2), the result of multiplying two near-synonyms is 1, the same as that of two identical words. For instance, "increase" and "Improve" are of the same meaning, if two sentences include "increase radiation" and "improve radiation", respectively, the multiplication result of the two words will be 1 after they are judged to be near-synonyms through the dictionary of near-synonyms.

IV.   CONCLUSION AND DISCUSSION

This paper proposes a segmentation algorithm based on special identifier and a new word recognition model based on search engine. Sentence-level similarity comparison model can evolve into three main similarity detection models: VSM, hamming distance and the longest common substring model, which obtains good effect on solving the paper plagiarism problem such as word deletion and addition, part copy and replacement of words by add order penalty factor and the process of handling near-synonyms. From the experiment result, it can be seen that the method proposed in this paper can evaluate paper similarity accurately. However, large-scale paper detection needs massive lexicon, data preprocessing and search library. Therefore, it needs more uniform standards in implementation and promotion to get better effects.

REFERENCES

[1] Manber, U. and G. Myers, Suffix arrays: A new method for on-line string searches. *SIAM Journal on Computing,* 1993. 22(5): p. 935-948.
[2] BAO Jun-Peng and SHEN Jun-Yi, A Survey on Natural Language Text Copy Detection. *Journal of Software,* 2003(10).
[3] SHI Yi, Variance clustering based outline identification algorithm for time series date. *Journal of Computer Application.*
[4] JIN Bo, SHI Yan-Jun and TENG Hong-Fei, Similarity algorithm of text based on semantic understanding. *Journal of DaLian University of Technology,* 2005(02).
[5] FENG Zhong-Hui, BAO Jun-Peng and SHEN Jun-Yi, Incremental Algorithm of Text Soft Clustering. *Journal of Xi'an Jiaotong Universiry,* 2007(04).
[6] HUANG Cheng-Hui, YIN Jian and HOU Fang, A Text Similarity Measurement Combining Word Semantic Information with IF-IDF Method. *Chinese Journal of Computers,* 2011(05).
[7] HUANG Cheng-Hui, YIN Jian and LU Ji-Yuan, An improved Retrieve Algorithm Incorporated Semantic Similarity for Lucene. *Journal of Zhongshan University (natural science edition),* 2011(02).
[8] WANG Xiu-Hong and JU Shi-Guang, Result merging method based on combined kernels for distributed information retrieval. *Journal on Communications,* 2011(04).

[9]   NIE Gui-Hua, Ontology-based Thesis Copy Detection System. *Computer Engineering,* 2009(06).

[10] ZHANG Huan-Jiong, WANG Guo-Sheng and ZHONG Yi-Xin, Text Similarity Computing Based on Hamming Distance. *Computer Engineering and Application,* 2001(19).

[11] Chen, Y. and M.C. Chen, Using chi-square statistics to measure similarities for text categorization. *Expert Systems with Applications,* 2011. 38(4): p. 3085-3090.

[12] S ánchez-Vega, F., et al., Determining and characterizing the reused text for plagiarism detection. *Expert Systems with Applications,* 2013. 40(5): p. 1804-1813.

[13] Sun, Yangguang and Cai, Zhihua, An improved character segmentation algorithm based on local adaptive thresholding technique for Chinese NvShu documents, *Journal of Networks,* v 9, n 6, p 1496-1501, 2014.

[14] Ksantini, Riadh1, Boufama, Boubaker, A new KSVM + KFD model for improved classification and face recognition. *Journal of Multimedia,* v 6, n 1, p 39-47, 2011.

[15] Madokoro, H, Sato, K., Facial expression spacial charts for describing dynamic diversity of facial expressions. *Journal of Multimedia,* v 7, n 4, p 314, 2012.

Uyghur Language Model with Graphic Structure

Miliwan Xuehelaiti [1, 2, 3*], Kai Liu [2], Wenbin Jiang [2], and Tuergen Yibulayin [1]

1. Xinjiang University Information Science and Technology institute Urumqi, Xinjiang, 830046, China
2. Institute of Computing Technology, University of Chinese Academy of Sciences Beijing 100190, China
3. Urumqi Administration of Industry and Commerce Urumqi, Xinjiang, 830002, China
*Corresponding author, Email: mihreban@126.com, {liukai, jiangwenbin}@ict.ac.cn

Abstract—**This paper describes a novel agglutinative language modeling strategy for Uyghur with graphic language model as structure. In graphic modeling language model, sentences are organized by morphemes as a directed graph, which is different from the linear structure in n-gram language models. The graphic language model is verified in two typical natural language processing application scenarios, morphological analysis and machine translation. The experiments show that the graphic language model achieves significant improvement in both morphological analysis and machine translation.**

Index Terms—**Uyghur; Graphic Language Model; Morphological Analyzing; Statistical Machine Translation**

I. INTRODUCTION

Language model is one of the most important models in NLP, it describes probabilities of sentences in natural language [1]. Many NLP tasks can be boiled down to the modeling of language model, such as transliteration, speech recognition, part of speech tagging and so on. One of most widely used language model is n-gram language model [3], which models words in sentences with local context environment in an linear way. The n-gram language model is simple and effective, and it have got excellent performance on Chinese, English and other languages with simple morphological form. Unlike Chinese and English, Uyghur is an agglutinative language. Agglutinative language is one kind of language which is widely used in North/South Korea, Japan, Mongolian, Turkey and other countries in Middle East Asia and other areas. Such languages differ from languages with simple morphological form (such as English and Chinese) in their sentence and word-formation [2], in which the composition of each word in agglutinative language follows different word-building rules according to simple observation:

Each word of agglutinative language is composed by a word-stem and any number of affixes, in where constrained relations exist between stem and affixes; and similar relations exist in stems of different words. The former rule lead to the data sparseness of word of agglutinative language, the latter rule makes it hard to seize the relation between stems, for there maybe some affixes between stems in different words.

According to the observations above, sentence of Uyghur with those relations can not be simply modeled as linear sequence. As a matter of fact, traditional n-gram language language model which models sentences as linear sequence can not obtain idea results on Uyghur. In this paper, we propose a novel graphic language model which can depict those relations more deeply. More specifically, our graphic language model models the generative relations between stem and affixes in a Uyghur word and the relations between stems in different words, which relations can hardly be modeled by traditional linear language models. In order to test the novel graphic language model, two language model needed natural language processing tasks (morphological analyzing and statistical machine translation (SMT)) are adopted to verify our graphic language model.

TABLE I. AN EXAMPLE OF UYGHUR WORD WITH MULTIPLE MORPHEMES. A WORD WITH DIFFERENT MORPHEMES WILL SHOW DIFFERENT MEANINGS AND EVEN BE A SHORT SENTENCE

word	stemming	meaning
zamaniwi	zamaniwi	modern
zamaniwilash	zamaniwi+lash	modernization
zamaniwilashtur	zamaniwi+lash+tur	modernization (it)
zamaniwilashtural	zamaniwi+lash+tur +al	can modernization (it)
zamaniwilashturalma	zamaniwi+lash+tur +al+ma	can not modernization (it)
zamaniwilashturalmam	zamaniwi+lash+tur +al+ma+m	can not modernization (it)?
zamaniwilashturalmamdu	zamaniwi+lash+tur +al+ma+m+du	can't they modernization (it)?

In addition to n-gram language model, there is much work devoted into language modeling. Some kind of structured language model aims at modeling the structures of language and overcoming the locality problem [10] and neural network is employed to improve the work [11]. But so far, there is not any work on the structure of agglutinative language. And there is much supervised work on morphological analyzing for each language respectively: Japanese [12], Arabic [13], and so on. Correspondingly, unsupervised ones (e.g. [14]) are also available. And morphological analyzing is proved to be an important task for other NLP task (e.g. SMT [16–18]). And morphological work can be done by properly modeling the relation of morphemes, where structured language modeling is an important task. For machine translation so far, most studies of agglutinative related machine translation are base on agglutinative to non-agglutinative translation, such as, for Turkish [15–17], Korean [18, 19] and others [20]. And there is also work on alignment between agglutinative language

and other languages in translation purpose [21, 22]. While there is less work on translation of non-agglutinative language to agglutinative language [23], where our well organized graphic language model will be a better choice.

In the experiments, both tasks show that graphic language model gets significant improvements compares to n-gram language model. In morphological analyzing, the accuracy gains 0.8% improvements due to the new style language model, while in SMT it gains more than 1.1 BLEU improvement. Furthermore, the graphic language model is simple, and the complexity of it is approximate to the n-gram language model.

The rest of paper is organized as follows: Section 2 describes the characteristics of Uyghur as an agglutinative language; Linear and proposed graphic language modeling methods for Uyghur are described in Section 3; And the methods of utilizing proposed graphic language model in two NLP tasks are shown in Section 4; Finally, we present the experiments of the two NLP tasks with graphic language model on Uyghur in Section 5 and conclude in Section 7.

II. UYGHUR AS AGGLUTINATIVE LANGUAGE

Uyghur is one of typical agglutinative languages, it is a Turkic language which is widely used in Western China by Uyghur people. Its words are made up of distinct morphemes by a linear sequence way, and each component of meaning is represented by its own morpheme. It has many characteristic, more specifically:

-Each word jointed with different affixes will show different meanings. Take a word in Uyghur as example, the word "xizmet" (job) will show different meanings when different morphemes followed with it: "xizmettin" (from work), "xizmetde" (with work) and so on.

- Each word can be jointed with multiple morphemes, and such a word can even be a short sentence. As it is shown in Table 1, the same word "zamaniwi" (modern) jointed with different morphemes convey different meanings. And much important information, such as, meanings of content words are conveyed by those morphemes.

In addition, morphemes of Uyghur fall into two categories: stem and affix. The first morpheme of each word is the stem of the word in Uyghur, and each word should have one and only one stem, which conveys the main semantic meaning of the word. As the example above, "xizmet"(job) and "zamaniwi" (modern)are stems. And all morphemes after stems are affixes, which convey minor semantic meanings or grammar information. Furthermore, all stems in a sentence form the skeleton of the sentence.

III. PROPOSED MODEL

Language model describes a word sequence $w_j^i(w_i, w_{i+1}, ..., w_{j-1}, w_j)$ by assigning the probability $p(w_j^i)$ to the sequence by means of certain probability distribution. And language model is widely used in many

NLP applications, such as speech recognizing, morphological analyzing, machine translation and so on.

Most language models regard the word sequence as a linear structure, and calculate the probability in a linear way. One of the most typical models is ngram language model, which assigns a given word's probability by means of its previous contiguous words. In n-gram language model, the probability of the sequence wij is assigned approximated as:

$$p(w_j^i) = \prod_k p(w_k \mid w_{k-1}^i) \approx \prod_k p(w_k \mid w_{k-1}^{k-n+1}) \quad (1)$$

where n is the n-gram size of the language model. And $p(w_k \mid w_{k-1}^{k-n+1})$ can be calculate from its frequency in corpus:

$$p(w_k \mid w_{k-1}^{k-n+1}) = \frac{\#(w_k^{k-n+1})}{\#(w_{k-1}^{k-n+1})} \quad (2)$$

where #(_) means the total count of the n-gram in the corpus. And in practical terms, the probability above needs some kind of smoothing, such as "add-one", Good-Turing, Kneser-Ney smoothing and so on [4].

A. Linear Modeling for Uyghur

Because of the characteristic of Uyghur, common linear modeling for will encounter some problems:

TABLE II.   THE LEXICAL STATISTICAL DATA FOR UYGHUR AND CHINESE. TOTAL WORD: THE TOTAL NUMBER OF WORDS EXISTED IN THE CORPUS; TOTAL FREQ: THE SUM OF ALL WORDS' FREQUENCIES; AVG FREQ: THE AVERAGE FREQUENCY OF EACH WORD

	uyghur	chinese
Total word	69563	43285
Total freq	1263861	1161801
Avg freq	18.17	26.84

-Data sparseness. For each word in Uyghur can be made of several morphemes, and the number of all probable words are astronomical, which can make serious data sparseness problem. For example, we count the lexical frequencies of both languages in parallel corpus (Table 2), from where we can see less frequency per word in Uyghur, in other words, more data spareness.

-Ignoring the relations in/between words. Traditional linear language model ignore to model relations between stems and relations inside the word. And a remedial measure for it is to model morphemes as basic units instead of words in n-gram model. This measure can describe stem-affix relations, but it still have weaker ability to describe the relations between stems.

B. Graphic Modeling for Uyghur

Since there are several drawbacks of linear modeling Uyghur, we propose a morpheme based directed graphic language model. And the directed graphic structure can better describe the characteristic of Uyghur.
As it is shown in Figure 1, we model an Uyghur sentence as directed graphic with morphemes as basic elements, where all stems are connected linearly in left-right order and all affixes are connected to previous affixes or stems. And it can be divided into two linear parts:
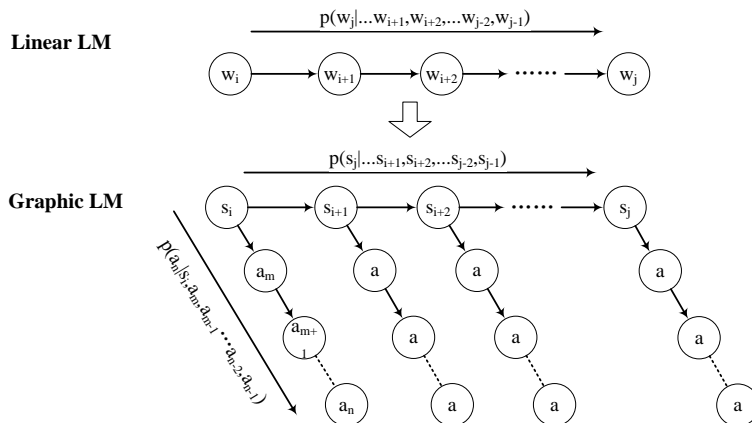
Figure 1. The structures of linear language model (Linear LM) and graphic language model (Graphic LM). Compared to linear language model, graphic language model describe the detail relation of stems and affixes inside/outside the words. And as it is shown in the figure, it can be separated into two parts of sub-models (stem-stem and stem-affix)

- stem-stem: this part assigns the probability to all stems, similar to n-gram model, the probability can be calculated approximately by assuming the probability of observing stem can be calculated in condition of preceding n stems:

$$p(s_i^1) = \prod_i p(s_i \mid s_{i-1}^1) \approx \prod_i p(s_i \mid s_{i-1}^{i-n+1}) \qquad (3)$$

where s_i denotes the stem of ith word.

- stem-affix: this part calculates all affixes probabilities by means of preceding stem and affixes in the same word:

$$p(a) = \prod_i \prod_j p(a_{i,j} \mid s_i, a_{i,j-1}^{i,1}) \qquad (4)$$

where $a_{i,j}$ denotes the jth affix in ith word in the sentence.

For the whole model, we combine both parts above together, and calculatesentences' probabilities as follows:

$$p(w_n^1) = \prod_i (p(s_i \mid s_{i-1}^{i-n+1}) \prod_j p(a_{i,j} \mid s_{i,} a_{i,j-1}^{i,1})) \quad (5)$$

where stems s and affixes a are obtained from morphological analyzing results of the sentence w_n^1. The training method of both parts of this model can refer to n-gram language model. In this way, we design a morpheme based directed graphic language model, which is supposed to be a better language model for Uyghur.

Training Firstly, we morphological analyze the training corpus into stemmed one. Secondly, according to the stemmed corpus, we obtain stem-stem and stemaffix corpora by removing affixes and splitting sentences by words respectively. Then, those corpora are utilized to train linear language model for stem-stem and stem-affix respectively.

Algorithm 1 Estimating Probability of Sequence
1: Stemmed Seq←MA(Seq)
2: Stem-Stem Seq←RemoveAffix(Stemmed Seq)
3: Stem-Affix Seq List←Split(Stemmed Seq)
4: Stem-Stem Prob←LM(Stem-Stem Seq, Stem-Stem Model)
5: **for each** Stem-Affix Seq ∈ Stem-Affix Seq List do

6: Stem-Affix Prob ←(Stem-Affix Seq,Stem-Affix Model)
7: **end for**
8: Graphic Prob ←Multiply(Stem-Stem Prob,Stem-Affix Prob)

Estimating Probability Algorithm 1 outlines the estimation procedure in its entirety. In line 1, we analyze the input sequence by morphological analyzing procedure MA(_) and obtain the stemmed corpus. The stemmed corpus is utilized to obtain stem-stem sequences by removing all affixes in the corpus in line 2. Correspondingly, in line 3 sentences are split according to words and then organized into stem-affix sequences. From line 4-7, we calculate the both parts' probabilities with corresponding sequences and sub models through procedure LM(_, _). And the final score of the model is combined by scores from sub-models in line 8.

IV. NLP APPLICATIONS

A. Morphological Analyzing

Morphological analyzing is one of the most important NLP tasks in agglutinative language [1]. The quality of morphological analyzing will affect other NLP tasks, which are based on morphological analyzing. There are three sub-task in morphological analyzing, including stemming, restoring the changed letter and POS tagging, in which we select the first sub-task stemming as the application to verify our graphic language model. Stemming is similar to segmentation, it splits each word into morphemes (Figure 2), including a stem and several following affixes. According to the characteristic of Uyghur, stemming needs the contexts of inside or outside the word, where language model is available and important.

Formally, we define a word sequence as w_j^i, which means it is a word sequence with words from position i to j in sentence. And each word can be segmented into, several morphemes m, which contain a single stem s and several following affixes a. In this task, we try to find the most probable morphological segmentation m_n^1 of sentence w_j^i.
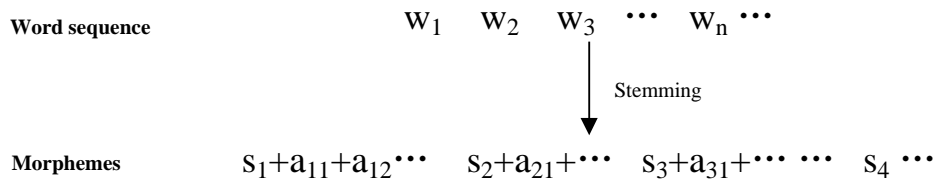
Word sequence $W_1 \quad W_2 \quad W_3 \quad \cdots \quad W_n \cdots$

Stemming

Morphemes $S_1+a_{11}+a_{12}\cdots \quad S_2+a_{21}+\cdots \quad S_3+a_{31}+\cdots \cdots \quad S_4 \cdots$

Figure 2. Stemming word sequence into morphemes, where the first morpheme in each word is stem and others are affixes

Word based:

Stem-Affix based:

Source sentence   Source sentence

Translation   Translation

$W_1 \quad W_2 \quad W_3 \quad W_4$ $S_1+a_{11}+a_{12} \; S_2+a_{21} \; S_3+a_{31}+a_{32} \; S_4+a_{41}$

Combination

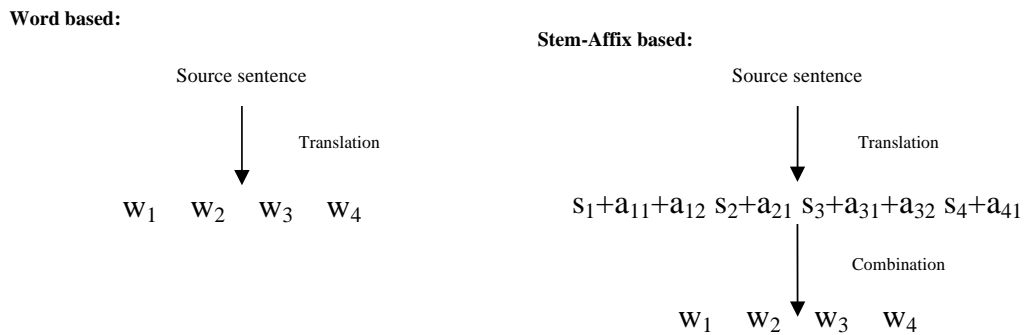$W_1 \quad W_2 \quad W_3 \quad W_4$

Figure 3. Translation from one language to Uyghur with different basic units. Left panel shows the translation from source language directly to target words, while the right panel shows the translation procedure that translate source language firstly into morphemes and then combine them into final words

With linear modeling, n-gram language model selects the morpheme sequence with the maximum language model probability:

$$\ell(w_n^1) = \arg\max_{m_{n'}^1} \prod_i p(m_i \mid m_{i-1}^{i-n+1}) \qquad (6)$$

where m denotes a morpheme, and $m_{n'}^1$ is the selected morpheme sequence of the sentence. And the $p(m_i \mid m_{i-1}^{i-n+1})$ means n-gram language model's probability of morpheme mi with context m_{i-1}^{i-n+1}.

With graphic modeling, correspondingly, we try to find the morpheme sequence with the maximum model probability with stems and affixes:

$$\ell(w_n^1) = \arg\max_{s,a} \prod_i P(s_i \mid s_{i-1}^0) \prod_{i,j} p(a_{i,j} \mid s_i, a_{i,j-1}^{i,0}) \qquad (7)$$

where si denotes the stem of the ith word, and ai,j denotes the jth affix of ith word. The first term of Formula 7 is the stem-stem part of our graphic model and the second term is the stem-affix part.

B. Machine Translation

Machine translation is one of the hardest problems in NLP. The performance of statistical machine translation is highly depended on the quality of the language model (cite), and it is a good task to verify the quality of language model. In this paper, we try to verify the effectiveness of graphic language model with Uyghur as target side.

Due to the characteristics of Uyghur, the application will be performed on two different SMT system with different granularities:

Word Based The SMT model is trained with words in Uyghur side as the basic translation unit, and this kind of SMT system has several characteristics:

-It has large-grained translation unit, which may suffer from data sparseness problem.

- Shorter sequences will be translated while we use word as basic translation unit.

- Word based translation system is free to recombination of morphemes into words.

Morpheme Based Correspondingly, the stemmed sentences are used to train the SMT model and stems and affixes are the basic translation unit here:

- Smaller-grained translation unit means less data sparseness problem.

- Longer sequences have to be translated while stems and affixes are chosen, which means higher translation complexity.

- Morpheme based translation system have to recombine stems and affixes into completed words.

V.   EXPERIMENTAL RESULTS

In this section, we verify our graphic language model through two applications. And there are three different types of language model will be utilized in the experiments:

- Word linear LM: the n-gram language model based on words, and it will be utilized in the application of SMT.

-Morphemes linear LM: morpheme based n-gram language model which will be utilized in both applications.

-Graphic LM: our graphic language model for Uyghur, and the order (n) of linear part is equal to corresponding comparison LM in experiments as default.

A. Morphological Analyzing

DataSet We make use of an annotated corpus Mega-words Corpus of Morphological Analysis of Uyghur, which is manually annotated by Xinjiang

multilingual key laboratory. And it contains about 67 thousands sentences, from which we select 5% as our testing set.

Training and Evaluation We train a 5-gram language model on morphemes and the linear part of our graphic model by SRI Language Modeling Toolkit [9] with Kneser-Ney smoothing. And we simply evaluate the stemming result by precision and recall of the morphemes.

TABLE III.   EXPERIMENT RESULTS ON MORPHOLOGICAL ANALYZING WITH DIFFERENT LANGUAGE MODEL

Model	P%	R%	F%
Morpheme linear LM	87.5	87.4	87.5
Graphic LM	88.0	88.6	88.3

Results As the results shown in Table 3, our graphic language model shows advantage on morphological analyzing compared to morpheme linear language model, where precision obtain an improvement of 0.5% and more than 1% improvement on recall.

B. Machine Translation

In this section, we compare our graphic language model with linear n-gram language model in SMT task. In SMT task, two different granularities are employed to verify the effectiveness of our graphic language model. One experiment is performed on word, while the other is performed on the results of morphological analyzing (stems and affixes).

DataSets For bilingual training data, we select Chinese-Uyghur corpus with 120 thousand parallel sentence pairs, which includes fifty thousand sentence pairs from corpus provided by CWMT 2011 evaluation task [5]. We obtain morphological result of the corpus by performing Uyghur morphological analyzer1 on the corpus. The parallel corpus's word alignments are obtained by running GIZA++ [6] on the corpus in both directions and applying "grow-diag-and" refinement.

Training and Evaluation We use the development set provided by CWMT 2011 evaluation task as our development set, and we organize 1000 sentence pairs as our own test set. The quality of translation is evaluated by the NIST BLEU-4 metric [7]. We make use of the standard MERT as the tuning algorithm to tune our cascaded translation model's parameters on development set.

Baselines and Our model we apply SRI Language Modeling Toolkit to train language models with modified Kneser-Ney smoothing on Uyghur side of the training corpus. The open source SMT decoder Moses [8] is selected as our baseline, which contains implementation of hierarchical phase model (Moses-chart). Correspondingly, our model is based on the same decoding system Moses, and train our graphic language model on the same corpus (training corpus).

Results The experiment result is shown in Table 4, which line 2-7 show the results of word based SMT model with different language model respectively. And line 9-10 give the results of morpheme based SMT model

with both linear language model and our graphic language model.

As the results shown, our graphic language model is significant better than those linear modeling language model. And both parts of our graphic language model (stem-stem, stem-affix) show their effectiveness on experiment, while the whole model shows better performance. Meanwhile, those improvements prove that it is reasonable to model Uyghur in stem-stem and stem-affix style, and this style of structure can describe some kinds characteristic of Uyghur.

TABLE IV.   EXPERIMENT RESULTS ON TEST SET. WE TEST TRANSLATION MODEL WITH DIFFERENT LANGUAGE MODEL RESPECTIVELY. WORD: WORD BASED N-GRAM LANGUAGE MODEL; STEM: STEM PART OF OUR GRAPHIC LANGUAGE MODEL; AFFIX: AFFIX PART OF OUR GRAPHIC LANGUAGE MODEL; MORPHEME: THE MORPHEME BASED LINEAR LANGUAGE MODEL. AND STEM+AFFIX IS OUR WHOLEGRAPHIC LANGUAGE MODEL. THE FOLLOWED NUMBER DENOTES THE ORDER OF THE LANGUAGE MODEL, FOR EXAMPLE, WORD5 MEANS A 5-GRAM WORD BASED LANGUAGE MODEL AND STEM3 MEANS WE TRAIN THE STEM PART WITH ORDER 3

Granularity	Language Model	BLEU%
Word based	word5	51.19
	word5+morpheme5	52.28
	word5+stem3	53.01 (+0.73)
	word5+stem5	53.18 (+0.90)
	stem3+affix3	53.18 (+0.90)
	stem5+affix5	53.44 (+1.16)
Morpheme based	morpheme5	54.26
	stem3+affix3	54.91 (+0.65)
	stem5+affix5	55.26 (+1.00)

VI.   CONCLUSION AND FUTURE WORK

In this paper, we model Uyghur with graphic structure on the basis of the characteristics of agglutinative language by observations. The novel language model can better describe Uyghur and remit data sparseness, where evidences are provided by the experiments of different NLP tasks. The experiment results show significant improvements on morphological analyzing and SMT tasks with 0.8F-score improvements and 1.1 BLEU improvements.

For future work, we will investigate other kinds of structures that can better model Uyghur (e.g. indirected graph or all connected graph) and involve more feature of Uyghur into our model, or directly model the language model as discriminative model.

REFERENCES

[1] Arianna Bisazza and Marcello Federico. 2009. Morphological pre-processing for Turkish to English statistical machine translation. *In Proceedings of IWSLT,* pp. 129C135.

[2] K. Oflazer, Two-level description of Turkish morphology, in *Literary and Linguistic Computing,* 1994, vol. 9, no. 2, pp. 137C148.

[3] Katz, Slava, Estimation of probabilities from sparse data for the language model component of a speech recognizer, Acoustics, *Speech and Signal Processing, IEEE Transactions on,* 1987, volume 35, 3, pp. 400-401

[4] Chen, Stanley F and Goodman, Joshua, An empirical study of smoothing techniques for language modeling, *Proceedings of the 34th annual meeting on Association for Computational Linguistics, Association for Computational Linguistics,* 1996, pp. 310-318

[5] Hongmei Zhao, Yajuan L ̈u, Guosheng Ben, Yun Huang, Qun Liu, The evaluation report of CWMT2011, *CWMT* 2011, 2011, pp. 261-180

[6] F.J. Och and H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics,* 29(1): 19C51.

[7] K. Papineni, S. Roukos, T. Ward, and W.J. Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. *In Proceedings of the 40th annual meeting on association for computational linguistics, pp. 311C318. Association for Computational Linguistics.*

[8] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B.Cowan, W. Shen, C. Moran, R. Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. *In Annual meeting-association for computational linguistics,* volume 45, page 2.

[9] A. Stolcke et al. 2002. Srilm-an extensible language modeling toolkit. *In Proceedings of the international conference on spoken language processing,* volume 2, pp. 901C904.

[10] Chelba, Ciprian and Jelinek, Frederick, Structured language modeling, *Computer Speech & Language,* 2000, volume 14, 4, pp. 283-332

[11] Emami, Ahmad and Jelinek, Frederick, A neural syntactic language model, *Machine learning,*2005, volume 60, 1-3, pp. 195-227, Springer

[12] Nagata, Masaaki, A stochastic Japanese morphological analyzer using a forward- DP backward-A* N-best search algorithm, *Proceedings of the 15th conference on Computational linguistics-Volume 1,* pp. 201–207, 1994, Association for Computational Linguistics

[13] Buckwalter, Tim, Buckwalter fArabicg Morphological Analyzer Version 1.0, 2002

[14] Creutz, Mathias and Lagus, Krista, Unsupervised morpheme segmentation and morphology induction from text corpora using Morfessor 1.0, 2005, *Helsinki University of Technology*

[15] Arianna Bisazza and Marcello Federico. 2009. Morphological pre-processing for Turkish to English statistical machine translation. *In Proceedings of IWSLT,* pp. 129C135.

[16] Sharon Goldwater and David McClosky. 2005. Improving statistical MT through morphological analysis. *In Proceedings of HLT-EMNLP,* pp. 676C683.

[17] Coskun Mermer and Murat Saraclar. 2011. Unsupervised Turkish morphological segmentation for statistical machine translation. *In Workshop of MT and Morphologically-rich Languages.*

[18] Young-Suk Lee. 2004. Morphological analysis for statistical machine translation. *In Proceedings of HLT-NAACL, Short Papers,* pp. 57C60.

[19] Minh-Thang Luong, Preslav Nakov, and Min-Yen Kan. 2010. A hybrid morphemeword representation for machine translation of morphologically rich languages. *In Proceedings of EMNLP,* pp. 148C157.

[20] Sami Virpioja, Jaakko J. Vayrynen, Mathias Creutz, and Markus Sadeniemi. 2007. Morphology-aware statistical machine translation based on morphs induced in an unsupervised manner. I*n Proceedings of MT SUMMIT,* pp. 491C498.

[21] Minh-Thang Luong and Min-Yen Kan. 2010. Enhancing morphological alignment for translating highly inflected languages. *In Proceedings of COLING,* pp. 743C751.

[22] Zhiyang Wang, Yajuan Lu, and Qun Liu. 2011. Multi-granularity word alignment and decoding for agglutinative language translation. *In Proceedings of MT SUMMIT,* pp. 360C367.

[23] Reyyan Yeniterzi and Kemal Oflazer. 2010. Syntaxto-morphology mapping in factored phrase-based statistical machine translation from English to Turkish. *In Proceedings of ACL,* pp. 454C464.

Miliwan Xuhelati was born in 1984. She is currently a doctoral candidate in Information science and Technology institute, Xinjiang University. Her main research interests include nature language process.

Kai Liu was born in 1985. He is currently a doctoral candidate in Institute of Computing Technology, University of Chinese Academy of Sciences. His main research interests include nature language process.

Wenbin Jiang was born in 1984. Ph. D. His main research interests include nature language process.

Tuergen Yibulayin was born in 1959. Ph. D. supervisor in Xinjiang University Information science and Technology institute.

Massive XML Data Mining in Cloud Computing Environment

Zhao Li

Department of Computer Science, Xinyang College of Agriculture and Forestry, Xinyang, China
Email: lizb5566@163.com

Abstract—**This paper concentrates on how to mine useful information from massive XML documents in cloud computing environment. The structure of the Cloud computing and the corresponding tree data model of a XML document are analyzed in advance. Afterwards, structure of the proposed XML data mining system is illustrated, which is made up of three layers, such as "Application layer", "Data processing layer", and "XML Data converting layer". In the XML Data converting layer, XML data are collected from databases and documents, and then the source data can be converted to XML file effectively. In the data processing layer, the process of data selection, cleaning and standardization for XML data set is implemented, moreover, a XML data set with higher degree of structure and rich semantics are obtained. In the application layer, "the results report module", "data query module" and "results analysis module" are included. Next, massive XML data mining algorithm is proposed. The main innovations of this algorithm lie in that 1) the structure of a XML document is represented as an unordered tree, 2) the sub-structures of a XML document are modeled as sub-trees, and XML trees are regarded as a forest which is made up of all the sub-trees. Experimental results show that the proposed method can effectively mine useful information from massive XML documents in cloud computing environment with high efficiency.**

Index Terms—**XML; Data Mining; Semi-Structured Data; Edit Distance**

I. INTRODUCTION

In recent years, Cloud computing has been a hot research field, because Cloud computing can provide flexible dynamic IT infrastructures, QoS guaranteed computing environments and so on. Furthermore, the concept of Cloud computing is defined as the computing mode, in which users can upload and download through servers in the Cloud freely [1].

Currently, Cloud computing is still evolving and there exists no widely accepted definition for it. One possible definition of Cloud computing is given as follows.

The technology of Cloud computing has shown the great potential in the field of IT application [2, 3, 4], software engineering et al. Programmers with creative ideas for novel Internet services no longer need the large capital outlays in hardware to utilize the proposed service or depend on humans. Furthermore, the applications of Cloud computing should not to consider the over-provisioning for a given service. In a word, using Cloud computing, the efficiency of information management can be promoted greatly [2-4].
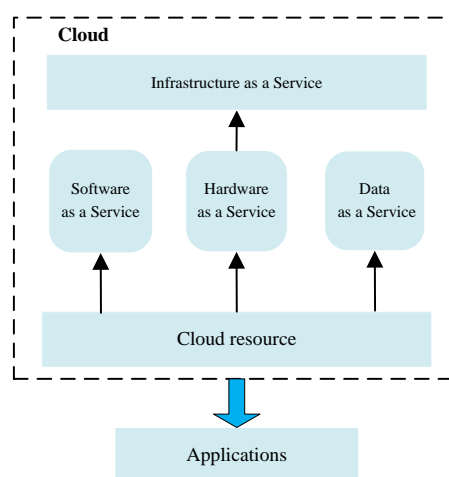

Figure 1.   Structure of the Cloud computing

As is shown in Fig. 1, Structure of the Cloud computing, which is made up of four main components, which are "Hardware as a Service" (HaaS), "Software as a Service" (SaaS), "Data as a Service" (DaaS) and "Infrastructure as a Service" (IaaS). Integrating all the above four parts, the cloud resource can be interacted with other applications.

In the Cloud computing model, Haas refers to hardware as a service, and it can meet most of the users' requirements. SaaS denotes the software as a service, which can reduce the users' workload of software maintenance, and the cost of software development. Particularly, DaaS means the data as a service, through which users could access the data in server easily. On the other hand, Cloud computing can give the infrastructure as a Service (IaaS) for users as well. Hence, users can access the suitable infrastructures with requirements of hardware configuration freely [1].

As is well known that, Extensible Markup Language (XML) refers to a markup language which defines a set of rules for encoding documents in a format that is both human-readable and machine-readable [5-7]. On the other hand, because XML file has become a popular format for the data storing and data exchanging, because XML has the features of the internal flexible semi-structures. Hence, it can be seen that XML can play an important role in data mining and intelligent information processing [9].
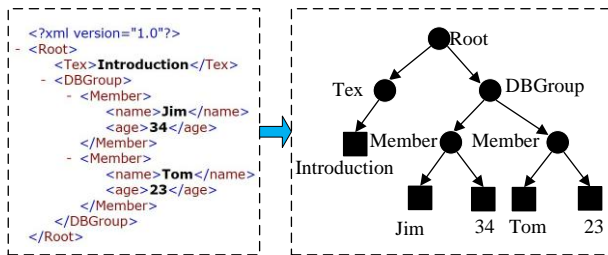
Figure 2.   A sample of XML file and the corresponding tree data model

XML documents can also be called the semi-structured data, which is traditionally modeled as Object Exchange Model (OEM), Many semi-structured data are represented by rooted trees with ordered nodes, which are called tree structured data [10]. For example, a rooted ordered tree represents the structure corresponding to the XML file (shown in Fig. 2).

Based on the above analysis, in this paper, we will focus on how to mining useful information from the massive XML data under the cloud computing environment.

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

1) The proposed approach is made up of three layers: "Application layer", "Data processing layer", and "XML Data converting layer". Massive XML data mining process can be implemented effectively using the proposed system.

2) In the XML mining processing, the structure of a XML document is represented as an unordered tree. Furthermore, the sub-structures of the XML document are represented as subtrees, and the XML tree is regarded as a forest which is made up of all the subtree.

3) Some new operations and new definitions are introduced in this proposed method.

The rest of the paper is organized as the following sections. Section 2 introduces the related works. Section 3 illustrates the proposed scheme for massive XML data mining under cloud computing environment. In section 4, experiments are conducted to make performance evaluation with comparison to other existing methods. Finally, we conclude the whole paper in section 5.

II.   RELATED WORKS

As XML is an important data representation, it has been widely used in data mining and intelligent computing. In the following section, the research about XML is given as follows.

Mazuran et al. presented a method of tree based association rules mining. In this method, mining rules provide approximate structure and the information content of the extensible markup language (XML) document content. Afterwards, the data obtained by the proposed method can be stored in the XML format as well. The mining of knowledge are utilized to provide the following information: 1) giving a brief idea including both the structure and the content of XML document and 2) providing fast approximate answers to users' queries [8].

Cokpinar et al. presented a rule mining framework which can give an efficient and effective method to solve the application of using positive and negative association rule computation running on the data streaming of XML format. Experimental results show that using XML to process data, this proposed framework is effective [9].

Jimenez et al. illustrated how scalable algorithms can be utilized to find frequent patterns from partially-ordered trees and then studied on the trade-offs which are generated in the proposed approach [10].

Romei et al. presented a system framework based on a general-purpose query language of which XML data are utilized. In the proposed system framework, raw data, mining models and human knowledge are described through XML data and then stored to the large-scale database. Afterwards, the proposed system focused on the problem of frequent pattern finding. Furthermore, the authors concentrate on how to exploit domain-dependent optimization in data extraction [11].

Zhang et al. proposed a novel system named XAR-Miner, which can mine useful information from XML documents with high efficiency. In this system, raw data in the XML data should be processed to transform to a XML Tree or to Multi-relational Databases. Particularly, this process needs to reduce the size of the XML document and low capacity of system memory to conduct the data mining process [12].

Chenier et al. illustrated a novel kind of broadcasting algorithm through association-rules in clients' request trends. In this paper, three kinds of algorithms are implemented, which are Exhaustive algorithm, Recursive algorithm, and Greedy algorithm. Finally, the proposed approach is tested and compared with other traditional methods [13].

Tang et al. studied on the methods of data mining algorithm used in the format of XML. The authors draw the conclusions that using the APIs provided in this paper can construct data mining application effectively. Furthermore, this paper showed that, apart from a few algorithms, the main contribution of SQL Server Data Mining can be used in OLE DB Data Mining process. Particularly, OLE DB for Data mining is an industrial standard which is designed Microsoft [14].

From the above related research works, it can be seen that XML is of great importance in massive data mining, especially in the Cloud computing environments.

As the page limitations of this paper, some other important research works about XML in recent years please refer to paper [16-23].

III.   PROPOSED SCHEME

A. Structure of the Proposed XML Data Mining System

In the massive XML data mining system, the XML language is utilized to compile the corresponding description documents for the database. Therefore, the use of such description files can reduce the database I/O burden to give an effective display mode for the original data sources. Particularly, our proposed algorithm can easily obtain the type of input data to produce output with a unified format. This process can facilitate the result

show and the use of other models. Furthermore, the proposed data mining approach can fast move between different computer systems, which aim to solve the problems in the XML data mining related applications.

As is shown in Fig. 3, the proposed system is made up of three layers, which are 1) Application layer, 2) Data processing layer, and 3) XML Data converting layer. The layer of XML data converting can collect use data from databases and documents, and then the source data can be converted to XML file effectively. For the data processing layer, the process of data selection, cleaning and standardization for XML data set is implemented, and then a XML data set with higher degree of structure and rich semantics could be obtained. In the application layer, the results report module, data query module and results analysis module are included in it.
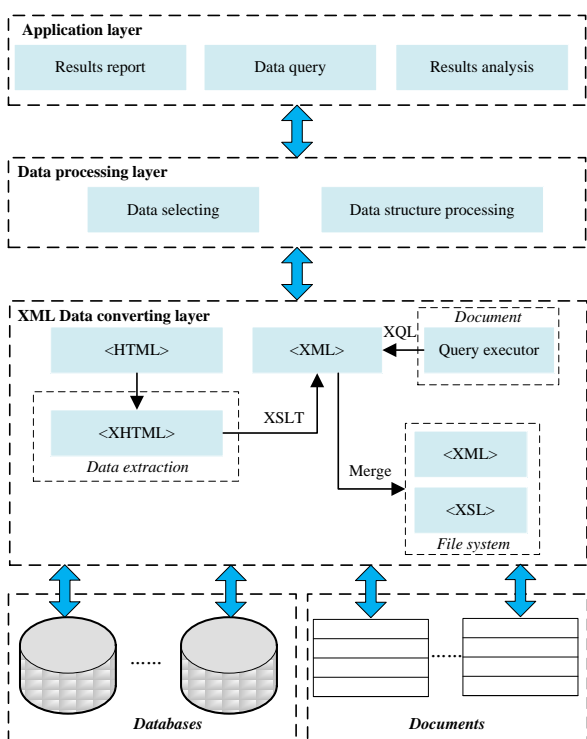

Figure 3.   Structure of the proposed XML data mining system

B. Massive XML Data Mining Algorithm

In this section, the structure of a XML document is represented as an unordered tree ($T = (N,E)$), N and E refer to the set of nodes and edges respectively. Afterwards, the sub-structures of the XML document could be modeled as subtrees. Particularly, the tree $t = (n,e)$ denotes a subtree of a XML tree T, and this relationship can be formatted as $t \mapsto T$. Based on the above analysis, the XML tree T can be regarded as a forest which is made up of all the subtree t, and the ancestor of the subtree t is represented as $\vec{t}$. Next, two based operation of the XML tree is defined as follows.

Definition 1 (Insert operation)

The insert operation can create a new node n, with node name and related value, as a child node of node m in an XML tree.

Definition 2 (Delete operation)

The delete operation refers to the inverse of the insertion one. It can remove the node n from a given XML tree.

Supposing that $\{t^i, t^{i+1}\}$ denotes the ith and $(i+1)^{\text{th}}$ historical states of the subtree t which is belonged to a XML tree T. $ED(t,i,i+1)$ denotes the edit distance of t from the ith state to the $(i+1)^{\text{th}}$ state. Hence, the degree of change for subtree t from state i to version $(i+1)$ is represented as follows.

$$DC(t,i,i+1) = \frac{ED(t,i,i+1)}{|t^i \cup t^{i+1}|} \qquad (1)$$

Supposing that $\{T^1, T^2, \cdots, T^n\}$ represent a sequence of n historical state of the XML tree structure T. Furthermore, ΔT_i denotes the set of subtrees which changes in the range of $[T^i, T^{i+1}]$. Hence, $\{\Delta T_1, \Delta T_2, \cdots, \Delta T_{n-1}\}$ represents the state of changed subtrees. For a set of subtrees ST, $ST = \{t_1, t_2, \cdots, t_m\}$, $\forall j \in [1,m], \exists i \in [1,n-1]$. Afterwards, the frequency of change for the set ST is defined as follows.

$$FC(ST) = \frac{1}{n-1} \cdot \sum_{i=1}^{n-1} V_i \qquad (2)$$

s.t. $\qquad V_i = \prod_{j=1}^{m} V_{ji} \qquad$ and

$$V_{ji} = \begin{cases} 1, & if\ DC(tj,i,i+1) \neq 0 \\ 0, & otherwise \end{cases} \quad j \in [1,m]$$

Supposing that $\{T^1, T^2, \cdots, T^n\}$ refers to a state of n historical states of a given XML tree. $\{\Delta T_1, \Delta T_2, \cdots, \Delta T_{n-1}\}$ denotes the state of changed subtrees. For a given minimum DC λ, the weight of the subtree set is defined as follows.

$$W(ST) = \frac{\sum_{i=1}^{n-1} D_i}{FC(ST) \cdot (n-1)} \qquad (3)$$

s.t. $\qquad D_i = \prod_{j=1}^{m} D_{ji} \qquad$ and

$$D_{ji} = \begin{cases} 1, & if\ DC(t_j,i,i+1) \geq \lambda \\ 0, & otherwise \end{cases} \quad j \in [1,m]$$

Next, supposing that $ST_i = \{t_1, t_2, \cdots, t_n\}$ refers a list of subtrees. $FT(ST_i)$ and $LT(ST_i)$ represent the first subtree and the last subtree which are belonged to ST_i respectively. Furthermore, the number of the weight counts which is needed to maintain for t is defined as follows.

$$N(WC) = \begin{cases} \prod\limits_{i=1}^{m-1} 2^{|ST_i|-1} \cdot 2^{|ST_m|}, \, if \, t \mapsto LT(ST_m) \\ \prod\limits_{i=1}^{m} 2^{|S_i|-1}, \, otherwise \end{cases} \quad (4)$$

Algorithm 1: Massive XML data mining algorithm
Input: Testing massive XML documents dataset
Output: XML data mining results

(1) Obtain the set of subtree sets $Q_{k-1} = \{q_1, q_2, \cdots, q_{k-1}\}$ from the input data

(2) **For** each element q_i, which is belonged to $Q_{k-1} = \{q_1^1, q_2^1, \cdots, q_{k-1}^1\}\}$ **do**

(3) **For** each element q_j, which is belonged to $Q_{k-1} = \{q_1^2, q_2^2, \cdots, q_{k-1}^2\}\}$ **do**

(4) **If** $(q_1^1 = q_1^2) \wedge (q_2^1 = q_2^2) \wedge \cdots \wedge (q_{k-1}^1 = q_{k-1}^2)$ then

(5) $c_k = \{q_1^1, q_2^1, \cdots, q_{k-1}^1, q_1^2, q_2^2, \cdots, q_{k-1}^2\}$

(6) **If** c_k contains at least one subset with the condition $FC < \lambda$ is satisfied **then**

(7) Delete the element c_k

(8) **Else** let $c_k \cdot N(FC)$ and $c_k \cdot N(WC)$ be equal to zero, and add c_k to the set C_k

(9) **End if**
(10) **End if**
(11) **End for**
(12) **End for**

IV. EXPERIMENTS

To make performance evaluation of the proposed algorithm, some existing XML data mining tools are compared, which are ebXMiner [24], VBUXMiner [25] and 2PXMiner [26]. Our proposed XML data mining tool is implemented with C++. Particularly, the computers in this experiment are set with Intel Corel i4 CPU, and the frequency of it is 2GHz. The memory we chosen is 8GB. Furthermore, the operating system is Windows 7.

The experimental dataset we used is named XMARK.DTD, which adopted DTDs to generate the rooted XML query trees. For this dataset, DTDs are converted into DTD trees in advance. Then, we construct a query database which includes 90,000 different queries. Afterwards, a number of queries are chosen randomly from the above process. In Table I, we show the characteristics of the XML query trees with a varying number of query trees from 10,000 to 90,000 for XMARK.

TABLE I.   BASIC SETTING OF THE XMARK.DTD DATASET

Query tree number	Max nodes	Min nodes	Max depth	Min depth
10,000	36	4	3	1
20,000	36	5	3	2
30,000	36	4	3	2
40,000	38	5	4	2
50,000	37	4	4	1
60,000	37	6	3	2
70,000	36	4	4	3
80,000	37	4	4	2
90,000	38	5	4	3

Next, we will introduce the settings of the XMARK.DTD dataset in Table I.

In experiment 1, the execution time and the number of the enumerated candidate subtrees of the above methods are compared under different number of XML query trees.
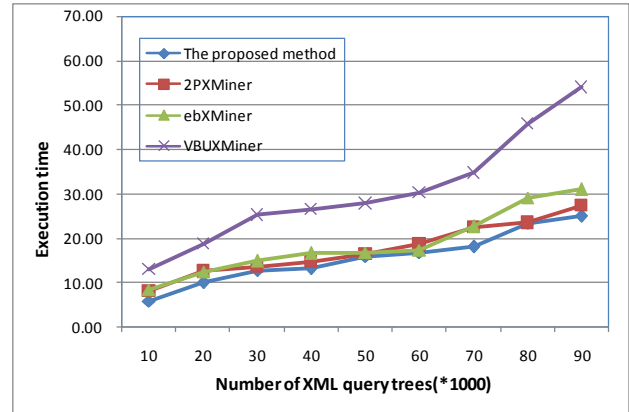
Figure 4.   Execution time when the number of XML query trees changing

From Fig. 4, we can see that the VBUXMiner need much longer execution time than other three miners. The reason lies in that the VBUXMiner is not suitable to be used in large-scale XML data mining.

Figure 5.   Enumerated candidate sub-trees when the number of XML query trees changing

Figure 6.   Execution time with minimum supports changing

Fig. 5 shows that the enumerated candidate sub-trees generated by the proposed method is larger than other

three methods. Furthermore, the execution time of the proposed method is smaller than other three miners. Therefore, it can be seen that the proposed method performs better than other methods both in performance and efficiency.

In experiment 2, as is shown in Fig. 6 and Fig. 7, we test the execution time and number of enumerated sub-trees of different methods under different minimum supports.

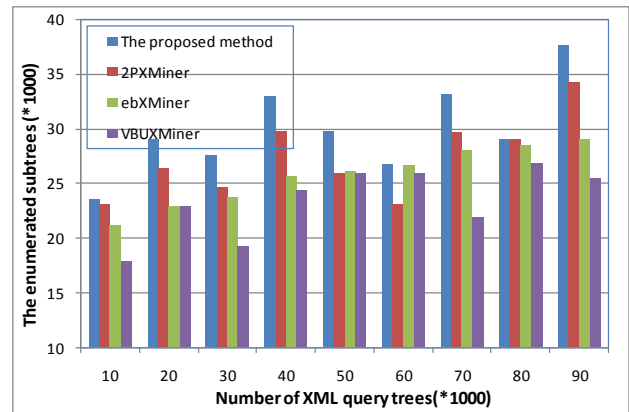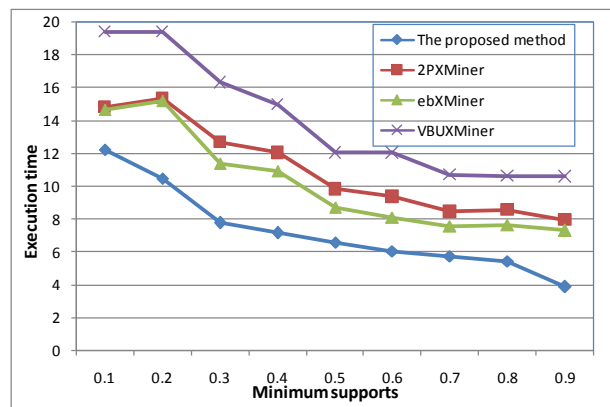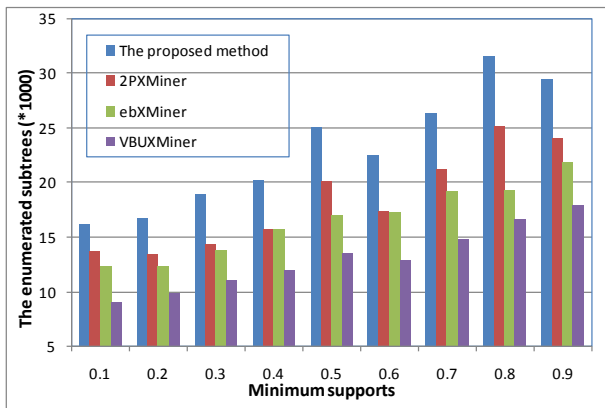Figure 7.   Enumerated candidate subtrees with the minimum supports changing

Combining the experimental results illustrated in Fig. 4-Fig. 7, the proposed performs better than other three methods both in execution time and in the number of enumerated candidate sub-trees.

Afterwards, we design experiment 3 to give a case study to illustrate the performance of the proposed algorithm in detail. The large-scale sample data are collected from the data of daily stocks transaction in Shanghai Stock Exchange Institute. We extract the data several times in the same way, and merge the extracted results each time into a single XML document. As is shown in Fig.8, the XML data mining results are illustrated. Particularly, the sample data used in this experiment are chosen in seven days.
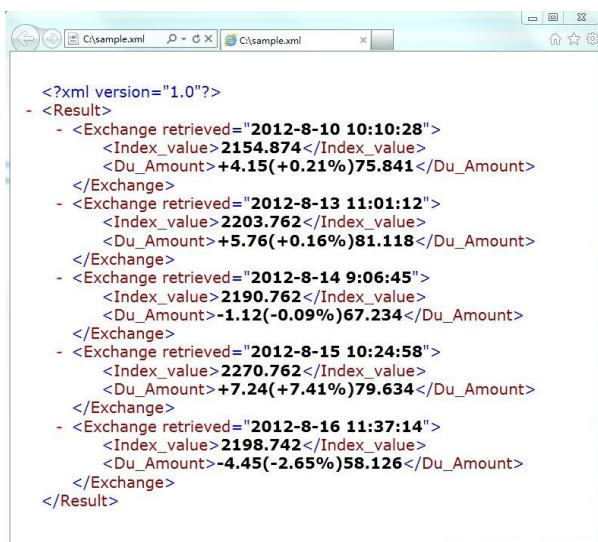

Figure 8.   XML data mining results from massive stock transaction data

The data mining results in Fig.8 show than the proposed method can effectively extract the "Exchange retrieved time", "Index value", "Transaction amount" from massive stock transaction data.

From the above experimental results, it can be seen that the proposed scheme is superior to other schemes. The main reasons lie in the following aspects:

1) VBUXMiner is implemented by two steps. In the first step, all queries are merged into a summary structure named "compressed global tree guide". In the second step, a bottom-up traversal scheme based on the compressed global tree guide is employed to generate frequent query patterns. To generate frequent query pattern trees rooted at node n in the compressed global tree guide, all frequent query pattern trees rooted firstly at the children of n should be generated, and then these frequent trees should be merged.

2) 2PXMiner refers to a two-pass mining algorithm that can detect frequent query patterns through retrieving database at most two times. Utilizing a transaction summary data structure and an enumeration tree, the upper bounds of the frequencies of the candidate patterns can be calculated. However, there are some limitations in this method. Firstly, this method has not considered the temporal issues which are important in capturing the user querying behavior. Secondly, this approach ignores the predicates in the query which are important in a real XML application.

3) ebXMiner outperforms VBUXMiner on the execution time. Both curves for VBUXMiner and ebXMiner increase as the nmber of XML query trees increases. Specifically, when the number of XML query trees is more than a given threshold, the improvement of ebXMiner is quite obvious. However, VBUXMiner does not consider the equivalent query trees in database, and results in more execution time are used for generating frequent query trees.

4) The proposed approach is made up of three layers: "Application layer", "Data processing layer", and "XML Data converting layer". The layer of XML data converting can collect use data from databases and documents, and then the source data can be converted to XML file effectively. The data processing layer has the performance of data selection, cleaning and standardization. In the application layer, the results report module, data query module and results analysis module are included in it.

5) In the proposed algorithm, the structure of a XML document is represented as an unordered tree. Particularly, the sub-structures of the XML document could be modeled as sub-trees, and the XML tree can be regarded as a forest which is made up of all the sub-trees.

6) "Insert operation" and "Delete operation" are defined for XML tree operating, which could enhance the performance of data retrieving in XML documents. The insert operation can create a new node with node name and related value, and this new node can be regarded as a child node of the given XML tree. The delete operation means the inverse of the insertion one. This operation can remove a node from the given XML tree.

V. CONCLUSIONS

In this paper, we present a novel massive XML data mining algorithm under the cloud computing environment. The structure of the proposed XML data mining system includes "Application layer", "Data processing layer", "XML Data converting layer". In the proposed massive XML data mining algorithm, the structure of a XML document is represented as an unordered tree, and the sub-structures of a XML document are modeled as sub-trees. Particularly, the XML tree is regarded as a forest which is made up of all the sub-trees.

REFERENCES

[1]   Wang Lizhe, von Laszewski Gregor, Younge Andrew, "Cloud Computing: a Perspective Study", *New Generation Computing,* 2010, 28(2) pp. 137-146

[2]   Armbrust Michael, Fox Armando, Griffith Rean, "A View of Cloud Computing", *Communications of the ACM,* 2010, 53(4) pp. 50-58

[3]   Hayes Brian, "Cloud computing", *Communications of the ACM,* 2008, 51(7) pp. 9-11

[4]   Wang Yulong, Shen Jiakun, "CloudProxy: A NAPT proxy for vulnerability scanners based on cloud computing, *Journal of Networks,* 2013, 8(3) pp. 607-615

[5]   Decker S, Melnik S, Van Harmelen F, "The semantic Web: The roles of XML and RDF", *IEEE Internet Computing,* 2000, 4(5) pp. 63-74

[6]   Arenas M, Libkin L, "A normal form for XML documents", *ACM Transactions On Database Systems,* 2004, 29(1) pp. 195-232

[7]   Schwentick Thomas, "Automata for XML - A survey, *Journal of Computer and System Sciences,* 2007, 73(3) pp. 289-315

[8]   Mazuran Mirjana, Quintarelli Elisa, Tanca Letizia, "Data Mining for XML Query-Answering Support", *IEEE Transactions on Knowledge and Data Engineering,* 2012, 24(8) pp. 1393-1407

[9]   Cokpinar Samet, Gundem Taflan Imre, "Positive and negative association rule mining on XML data streams in database as a service concept", E*xpert Systems with Applications,* 2012, 39(8) pp. 7503-7511

[10] Jimenez Aida, Berzal Fernando, Cubero Juan-Carlos, "Mining frequent patterns from XML data: Efficient algorithms and design trade-offs", *Expert Systems with Applications,* 2012, 39(1) pp. 1134-1140

[11] Romei Andrea, Turini Franco, "XML data mining", *Software-practice & Experience,* 2010, 40(2) pp. 101-130

[12] Zhang Ji, Liu Han, Ling Tok Wang, "A framework for efficient association rule mining in XML data, *Journal of Database Management,* 2006, 17(3) pp. 19-40

[13] Chenier Cameron, Jun J. James, Zhang Jason, "Association-rules mining based broadcasting approach for XML data", *Lecture Notes in Computer Science,* 2006, 4243 pp. 207-216

[14] Tang ZH, Maclennan J, Kim PP, "Building data mining solutions with OLE DB for DM and XML for analysis", *Sigmod Record,* 2005, 34(2) pp. 80-85

[15] Zhang J, Ling TW, Bruckner RM, "On efficient and effective association rule mining from XML data", *Lecture Notes in Computer Science,* 2004, 3180 pp. 497-507

[16] Zhang WS, Liu DX, Zhang JP, "A novel method for mining frequent subtrees from XML data, *Lecture Notes in Computer Science,* 2004, 3177 pp. 300-305

[17] Fang CH, Zhang YX, Xia DL, "A XML-based data communication solution for program mining", *Lecture Notes in Computer Science,* 2003, 2690 pp. 569-575

[18] Chonka Ashley, Xiang Yang, Zhou Wanlei, "Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks", *Journal of Network and Computer Applications,* 2011, 34(4) pp. 1097-1107

[19] Yang Donghua, Feng Yuqiang, Yuan Ye, "Ad-hoc aggregate query processing algorithms based on bit-store for query intensive applications in cloud computing", *Future Generation Computer Systems-the International Journal of Grid Computing and Escience,* 2013, 29(7) pp. 1725-1735

[20] Costa Rostand, Brasileiro Francisco, Lemos Guido, "Analyzing the impact of elasticity on the profit of cloud computing providers, *Future Generation Computer Systems-the International Journal of Grid Computing And Escience,* 2013, 29(7) pp. 1777-1785

[21] Cala Jacek, Hiden Hugo, Woodman Simon, "Cloud computing for fast prediction of chemical activity", *Future Generation Computer Systems-the International Journal of Grid Computing and Escience,* 2013, 29(7) pp. 1860-1869

[22] Moreno-Vozmediano Rafael, Montero Ruben S., Llorente Ignacio M, "Key Challenges in Cloud Computing Enabling the Future Internet of Services", *IEEE Internet Computing,* 2013, 17(4) pp. 18-25

[23] Yang Chaowei, Xu Yan, Nebert Douglas, "Redefining the possibility of digital Earth and geosciences with spatial cloud computing", *International Journal of Digital Earth,* 2013, 6(4) pp. 297-312

[24] Bei Y.-J., Chen G., Dong J.-X., Chen K., Bottom-up mining of XML query patterns to improve XML querying, *Journal of Zhejiang University,* 2008, 9(6) pp. 744-757

[25] Bei Y., Chen G., Shou L., Li X., Dong J., Bottom-up discovery of frequent rooted unordered subtrees. *Information Sciences,* 2009, 179 pp. 70-88

[26] Yang L. H., Lee M. L., Hsu W., Huang D., Wong L., Efficient mining of frequent XML query patterns with repeating-siblings. *Information and Software Technology,* 2008, 50 pp. 375-389

Face Recognition Based on Wavelet Transform and Regional Directional Weighted Local Binary Pattern

Wu Fengxiang

North China Career Academy of Water Resources, Henan Zhengzhou, China
Email: yancaifeng2002@163.com

Abstract—**With the development of information technology, face recognition technology has been continuously developed. This technology has attracted the attention of many researchers, including institutions and production enterprises. Face recognition technology has become a relatively independent application technology area in various social services. This paper presents a face recognition algorithm based on wavelet transform and regional directional weighted local binary pattern. First of all, this algorithm puts forward a new basis for a face recognition, namely the level of detailed components of face images containing valid facial texture details, and the recognition rate is better than that of the vertical component information and diagonal component information. This is called Horizontal Component Prior Principle(HCPP). According to HCPP, the original image is decomposed with wavelet transformation. The algorithm extracts the scale and level of detailed components. To improve the original LBP operator, it presents the regional directional weighted local binary pattern (RDW-LBP). Using the RDW-LBP, it can calculate the histogram of scale components and detailed components decomposed by wavelet. The histogram feature vector of face image can be got with the different weighted sub-regions. The feature vector can be matched with Chi-Square distance. This approach further enhances the ability to extract face direction information effectively.**

Index Terms—**Face Recognition; Feature Extraction; Regional Directional Weighted Local Binary Pattern**

I. Introduction

Face image is a kind of biological information which can achieve through non-contact. Personal identification has been a research hotspot in the field of pattern recognition and artificial intelligence. Over the years, researchers from all over the world have done many researches, but due to variability in the practical application environment such as the non uniqueness and non rigidness, face images make a certain distance between all kinds of recognition algorithm performance and practical business application requirements[1]. Therefore, further improving the recognition rate and robustness of face recognition algorithm become a main goal of face recognition research.

Pre-processing part of face image consists of four processes, namely the human face image normalization, scaling, illumination compensation and mask processing.

1. Face image normalization. According to the result of eye location, it can normalize the face database image to reduce the influence of deflection and scale change on recognition.

2. Scaling. Face image scaling can be realized by using bilinear interpolation algorithm for the calculation of wavelet decomposition. In this paper, a unified image scaling is 128 by 128.

3. Illumination compensation. Illumination change is one of the most important factors affecting the recognition rate. The problem of facial shadow caused by the uneven and excessive light will cause the recognition rate fall sharply. Adequate illumination compensation is particularly important to improve the recognition rate. Method to solve the illumination problem is divided into the following two categories. The first way is to start from the point of view of image enhancement, by the adjustment of brightness, contrast and facial image histogram distribution, to reduce the effect of illumination. Commonly used methods are histogram equalization, Gamma correction, logarithmic transformation, frequency filtering [9]. The second category is to recover the reflection component by splitting light components in the face image to effectively eliminate the effects of lighting [10-12]. In this paper, it uses the histogram equalization and LOG-DCT and Gamma correction algorithm for illumination preprocessing.

4. Mask processing. Mainly for compression and recognition of irrelevant or relatively changed areas, it uses mask processing which is assigned according to location algorithm or test library.

Basic Local Binary Pattern (LBP) was firstly put forward by Ojala [2] which was applied to the lines principle analysis and image retrieval. The late extension for unified mode circular neighborhood LBP descriptor [3] has been proposed to further reduce the dimension of the texture and to improve the ability of its portrayal of texture. TimoAhonen [4] who firstly proposed LBP operator applied to face recognition expanded a new research idea in this field in 2004.

Without training, LBP algorithm has the strong promotion and classification ability. But for all the problems, it uses a single transformation which is not conducive to seek the classification hyperplane. At the

same time, extracting histogram vector in more images will inevitably lead to high vector dimensions which will make the computational complexity increases. If images are less, it will lose statistical significance. Otherwise, the robustness and illumination problems of LBP algorithm also limit the application.

By using basic LBP operator based on local texture feature, TimoAhonen [5] introduced the idea of weighted sub-block on the basis of face image which further strengthened a lot of detailed characteristics of the eye and mouth area to affect recognition results. The method on FERET face database achieved good recognition performance. ShengcaiLiao [6] proposed a method of multi-scale Block LBP, which separated image into blocks on the basis of calculating the first level scale image LBP map and used average grey value of subblocks to generate the second level scale image. Finally the multistage scale LBP histogram can be connected as a feature vector. The LBP feature histogram calculated by this way can contain both the micro and macro structure characteristics of the image, so the algorithm is more robust.

In order to improve the LBP operator's ability to depict texture directionality, Mr Zhang proposed a method of combining LBP operator with Gabor wavelet for face recognition algorithm. First of all, this algorithm used multi-resolution Gabor wavelet filter for the normalized face image and extracted multiple Gabor amplitude domain mapping (Gabor Magnitude Map, GMM) corresponding to different directions and different scales, and then used LBP operator to calculate the local neighborhood relation on each GMM model. Combining with the Fisher discriminant classifier, the weighted matching HSLGBP binary was used for classification. Finally the algorithm on FERET face database also achieved good recognition performance. Wang wei extracted LBP feature spectra of two-stage low-frequency component by discrete wavelet decomposition, the test also achieved good recognition rate.

Xiaoyang Tan and Bill Triggs [7] [8] proposed a scalable LBP texture description operator –Local Ternary Pattern, (LTP). Threshold function of the LBP operator was modified by LTP operator. A original certain threshold value was improved to a set of threshold value intervals. And binary texture pattern of LBP was improved to the ternary texture patterns. At the same time, in order to reduce the complexity of the algorithm, the standard LTP texture model was divided into ULBP (Upper LBP) and LLBP (Lower LBP)-two LBP texture pattern for processing. The symmetry and noise threshold of LTP texture pattern can effectively eliminate the noise which improved the defects of LBP mode being sensitive to noise to improve the recognition rate. However, LTP has higher characteristic dimension problem, which makes the LTP feature need to take up more memory in storage and classification matching and makes the algorithm complex. It has seriously affected the further application of LTP.

So far, from previous research the following conclusions can be summarized. LBP operator describes the details of image, it has high distinguish ability to similar face images. But because LBP descriptor is small, it is more sensitive to noise and illumination. At the same time, the previous extension and improvement of LBP operator is proposed on the basis of the neighborhood pixel, they ignored the details of a microscopic scale and different direction which will affect recognition rate.

In the next section, we study process of face image preprocessing. In Section 3 we propose an improved optimization algorithm. In Section 4, we test the performance of proposed scheme and compare it with other face recognition algorithms. In Section 5 we conclude the paper and give some remarks.

II. PROCESS OF FACE IMAGE PREPROCESSING

Light, deflection, size, background and other objective factors are the enormous challenges for face recognition algorithms to improve the recognition rate. These factors must have certain robustness which is one of the main criteria for a practical recognition system. The effective pretreatment is a major method to reduce the influence of external factors. Pre-processing part of face image consists of four processes, namely the human face image normalization, scaling, illumination compensation and mask processing.

A series of different resolution sub-images by wavelet transform are obtained to form the original face image. Different sub images correspond to different frequency. The same level of the high frequency in sub-images reflects the decomposed images in different directions on the details. Low-frequency sub-images contain the main information in decomposed images. The eyes and the mouth have the main effect for recognition than other areas in a face image. After wavelet decomposition in face image, 3 detailed components can be obtained, and the horizontal detailed component contains the more features than other components in facial recognition. That's the reason why the horizontal component is taken as the priority principle. 2D-Gabor function is defined as (1).

$$\psi_{u,v}(x,y) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2(x^2+y^2)}{2\sigma^2}\right) \cdot \left[\exp\left(ik\left(\frac{x}{y}\right)\right) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \quad (1)$$

$$k = \begin{pmatrix} k_x \\ k_y \end{pmatrix} = \begin{pmatrix} k_v \cos\varphi_u \\ k_v \sin\varphi_u \end{pmatrix}, \quad k_v = 2^{-\frac{y+2}{2}\pi}, \quad \varphi_u = \frac{u\pi}{K}.$$

x, y are Coordinate values of pixels. u is direction of Gabor wavelet. K is the number of total direction. v is scale factor of Gabor wavelet. $\frac{k}{\sigma}$ determines the size of Gaussian window. The convolution of $I(x,y)$ and $\psi_{u,v}(x,y)$ is (2).

$$M(u,v,x,y) = \psi_{u,v}(x,y) \otimes I(x,y) = \iint \psi_{u,v}(x,y) \cdot I(x,y)dxdy \quad (2)$$

III.   AN IMPROVED OPTIMIZATION ALGORITHM

A kind of algorithm based on wavelet transform and regional directional weighted local binary pattern is proposed for face recognition. First, the algorithm puts forward a new aspect for face recognition basis that the horizontal component details of the face image contain more effective facial textures than vertical component ones and also the level component details do more contribution than vertical component ones for recognition rate. Here is called a horizontal component prior principle (HCPP). According to the HCPP, algorithm uses wavelet transformation to the original image decomposition to extract the scale of the components and horizontal detail. To improve the original LBP operator, the regional directional weighted local binary pattern (RDW-LBP) is proposed. After the wavelet decomposition of the scale component and level detail component, its DW-LBP histogram can be calculated, and different sub areas are weighted in marco meaning, a character vector of RDW - LBP histogram is obtained corresponding to the face image. Finally, the Chi-Square distance is used for sequence histogram matching. By improving the primitive image calculation method of the LBP, RDW-LBP intensifies the ability to extract the facial texture information in effective direction.

LBP is based on the center pixel gray value and the local neighborhood pixel gray value to generate decimal code between two values and then calculate the histogram of LBP image to describe the texture details. The LBP operator basically consists of a 3 by 3 matrix which contains 9 pixel gray values. Suppose that gray value of the center pixel is g_c, 8 pixels around the center respectively are $g_0 \sim g_7$ (Fig. 1). The formula for the calculation of the basic LBP is Fig. 2.

$$LBP(g_c, g_i) = \sum_{k=0}^{7} 2^k S(g_i - g_c) \qquad (3)$$

$$S(g_i - g_c \mid i = 2,6) = \begin{cases} 1, g_i - g_c > g_0 \\ 0, g_i - g_c < g_0 \end{cases}$$

$$S(g_i - g_c \mid i = 0,4) = \begin{cases} 1, g_i - g_c > g_1 \\ 0, g_i - g_c < g_1 \end{cases} \qquad (4)$$

$$S(g_i - g_c \mid i = 1,3,5,7) = \begin{cases} 1, g_i - g_c > g_2 \\ 0, g_i - g_c < g_2 \end{cases}$$

It can be promoted to coding pattern of circular neighborhood. Due to pixels symmetry of circular neighborhood, we can calculate threshold setting of $[0°, 90°]$ interval, then by symmetry transformation, we can get threshold distribution of the whole circle domain. It is supposed that there are P number of sample pixel point, $P = 2^k$. A circle is divided into four intervals and take the IV interval for threshold sorting analysis. The point which is closer to number $g_{4 \times 2^{k-2}}$, the corresponding binary threshold $\delta_i, i = 1, 2, \ldots, k$ is less. The point which is closer to number $g_{3\frac{1}{2} \times 2^{k-2} - 1}$, the corresponding binary threshold $\delta_i, i = 1, 2, \ldots, k$ is bigger and $\delta_0 \leq \delta_1 \ldots \leq \delta_k$.

In order to improve the rotation invariant of LBP operator, it is a good method to change the square neighborhood to circular neighborhood with arbitrary radius.
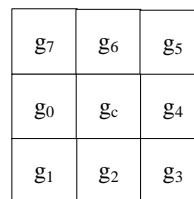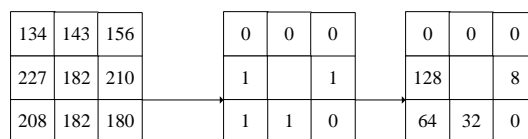
g_7	g_6	g_5
g_0	g_c	g_4
g_1	g_2	g_3

Figure 1.   LBP operator

134	143	156
227	182	210
208	182	180

0	0	0
1		1
1	1	0

0	0	0
128		8
64	32	0

Binary:11101000
LBP=128+64+32+8=232

Figure 2.   Calculation of LBP

So coding mode can be expressed as (5) to (7) and $BI(\cdot)$ represents bilinear interpolation calculation.

$$\{g_i \mid i = 0, 1, \ldots, k\} = BI(g_c, R, \theta) \qquad (5)$$

$$DW - LBP(g_c) = \sum_{k=0}^{P-1} 2^k S(g_i - g_c) \qquad (6)$$

$$S(g_i - g_c \mid i = 0, 1, \ldots, P-1)$$
$$= \begin{cases} 1, & g_i - g_c \geq \delta_i \\ 0, & g_i - g_c < \delta_i \end{cases} \qquad (7)$$

From the definition of the LBP operator, it can be thought as a fusion of statistical method and structure analysis method for texture feature extraction. Each pixel in LBP will find a best matching code generated by original texture. LBP can overcome the deficiency of traditional statistical or structural method and has strong ability to describe the texture feature.

In this paper, binary local model RDW - LBP is proposed based on texture according to the different weighted directions. RDW - LBP thinks that the special texture of the face image, also is in line with the HCPP. So the local neighborhood pixels should not be assigned with equal weighting, and should be treated differently. Because horizontal direction information is more conducive to identify, horizontal direction should be given greater weight.

Figure 3.   The original image and output with LBP

Take the eight neighborhood pixels of 3 by 3 matrix for example (Figure 1), g_0 and g_4 are located as the level

adjacent pixels to the center pixel g_c. Their gray level difference with the center pixel represents brightness change of the local texture in the horizontal direction, thus it has maximum weight. g_2 and g_6 is directly below and above on the center pixel, and the weight takes second place. g_1, g_3, g_5 and g_7 have diagonal location with the center pixel, and they have minimum weight.

According to the algorithm, the weight is updated as shown in (8).

$$Wi, c = \exp(-\|g_i - g_c\|^2 / t) \qquad (8)$$

So the vector can be calculated as (9):

$$Q_i = \sum_{k=0}^{N} 2^k S(g_i - g_c) \cdot W_{i,c} \qquad (9)$$

N represents the number of the matrix. Based on RDW - LBP, different areas of the face image can be made with different weight. Fig. 3 shows that the 64 x 64 resolution of face images are divided into different number of blocks, as well as the weighting of the sub-block. According to the following situation, eye block has the maximum weight (dark grey), mouth weight (light gray) is smaller than eye block weight, cheeks and forehead weights (white) are smaller than mouth weight, and the weight of the rest parts is 0, which represents area not be included in the calculation. In practice, the weight can be adjusted according to the result of face positioning.

Here the function can be shown as follows:

$$P^{(n-1)} = (Q_1, Q_2, \ldots Q_{n-1}) \qquad (10)$$

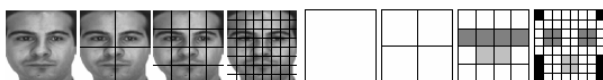$$R^{(n-1)} = \left[P^{(n-1)}\right]^T D_w P^{(n-1)} \qquad (11)$$

Figure 4.   Different areas of the face image with different weight

D_w is a matrix with w dimension. When the best facial feature vector is extracted, it is needed to design an effective classifier to classify feature vector. It plays an important role of decision-making mechanism for classifier. If the classifier performance is good, it can achieve ideal classification results even though sometimes the extracted features are not good enough. On the contrary, if the classifier design is not good, it may not be able to achieve better classification results even the extracted characteristics are very good. Here are several kinds of classifiers.

The Euclidean Distance is usually a distance definition. It is the true distance between two points in the m dimensional space. The smaller the distance is, the more similar two histograms are.

$$d(H_1, H_2) = \sqrt{\left(\sum_i (H_1(i) - H_2(i))^2\right)} \qquad (12)$$

The Chi-Square. For chi-square, the higher matching degree is, the smaller the distance is, and the chi-square value is lower.

$$d(H_1, H_2) = \sum_i \frac{(H_1(i) + H_2(i))^2}{H_1(i) + H_2(i)} \qquad (13)$$

The Bhattacharyya Distance is used to measure the correlation of two groups of feature histogram, which is commonly used for classifier algorithm.

$$d(H_1, H_2) = \sqrt{1 - \sum_i \frac{\sqrt{H_1(i) \cdot H_2(i)}}{\sum_i H_1(i) \sum_i H_2(i)}} \qquad (14)$$

In this paper, the Chi – Square is used to classify extracted characteristic vector. Assume that using n-tier wavelet decomposition, detail coefficients can be obtained after the multistage scale decomposition. With the multistage scale decomposition of coefficient, each sub- figure is divided into m by m sub-regions. The sub-regions are separately calculated into RDW-LBP histogram, and all the RDW-LBP histograms are concatenated into a sequence of facial features.

IV. EXPERIMENT AND ANALYSIS

In order to test the RDW-LBP algorithm, this paper selects the most commonly used two people face databases to test face recognition, which are AR face database and ORL face database respectively. In order to eliminate gender difference in recognition rate, this article from the AR face database randomly selects 50 people of the 70 men and 50 women face images for training and testing. Because this article focuses on testing recognition algorithm efficiency in shining environment, we select 12 images without shelter in each 24 pictures, 6 pieces are for training, and 6 copies are for testing. In ORL face database, it selects the volunteers in the experiment with odd number 5 images for training and even number 5 images for testing. TABLE I is recognition rate with three wavelet decomposition.

Based on horizontal component priority principle, Level details of face image do the largest contribution for recognition rate, which is greater than the vertical components and the diagonal components.

Because the different wavelet image decomposition can produce different results. In order to test the wavelet type influence on recognition rate, this paper compares the Haar wavelet, Daubechies wavelet and wavelet Sym-4 in AR and ORL face databases. TABLE I shows the wavelet's impact on the final recognition rate. From each group of LL+LH, it can be seen in the column data, Haar wavelet shows the optimal effect. Its main reason is that, although the three low frequency component of the decomposed wavelet level LL, Haar wavelet information retention is incomplete, but it retained the most abundant high frequency LH details accordingly. For face recognition, different wavelet base for the low-frequency component are not obvious, the high frequency components provide the more detailed information, and the more conducive to improve recognition rate. The ultimate recognition rate of using Haar wavelet is optimal.

TABLE I.   RECOGNITION RATE WITH THE THREE WAVELET DECOMPOSITION

	Haar (%)		Daubechies (%)		Sym-4 (%)	
	LL+LH	LL	LL+LH	LL	LL+LH	LL
AR	97.2	96.2	96.5	94.1	96.3	94.4
ORL	99.2	98.7	98.7	98.3	98.6	97.8

Wavelet is an important tool for multi-scale analysis. Different decomposition layers can extract the details in different scales and can show different facial features. But because the LBP operator is the local texture feature, excessive decomposition layers may make the number of pixels too little in matrix, which will make LBP operator more sensitive to noise. Finally, it will affect the recognition. Therefore, this paper tests the effect of different wavelet decomposition layers on the final recognition rate. The Haar wavelet which has been tested with high recognition rate above is selected to test the recognition rate with different decomposition layers. TABLE II is recognition rate with different decomposition layers and TABLE III is recognition rate of different algorithms in the AR face recognition. TABLE II shows that the increase of Haar wavelet decomposition level can improve the recognition rate. But it is not the more the better. From the table II, it shows the recognition rate with 3 levels Haar wavelet decomposition is less than the recognition rate with 2 levels Haar wavelet decomposition. Also, the 3 level Haar wavelet decomposition increases the complexity and leads to the decrease of the calculation speed. So 2 levels Haar wavelet decomposition is used to extract features.

TABLE II.   RECOGNITION RATE WITH DIFFERENT DECOMPOSITION LAYERS

	1 level	2 levels	3 levels
AR	96.5%	96.67%	96.5%
ORL	98.4%	98.7%	98.3%

This article uses the Haar wavelet and RDW - LBP for human face feature extraction and Chi - Square for recognition test. On AR and ORL face databases, it is helpful to select the PCA, LBP and RDW-LBP recognition algorithm for recognition test and analysis. 500 pictures in AR face database are selected and added Gaussian noise the mean of which is 0 and the variance is 0.05, and the correct identified pieces are 489. The results increases by 4.1% and 1.8% respectively compared with PCA, LBP recognition rate (TBALE III); In ORL face database, 250 pictures are used for test and added Gaussian noise the mean of which is 0 and the variance is 0.05, the correct identified pieces are 243 pieces. The results increases by 5.5% and 3% respectively compared with PCA and LBP recognition rate (TABLE IV). Experimental data shows that RDW - LBP has strong face texture feature describing ability and robustness of facial expression.

As can be seen from Fig. 5 and Fig. 6, the red line represents the RDW-LBP algorithm, the blue line represents LBP algorithm. Number 1 represents matrix of 2*2, 2 represents matrix of 3*2, 3 represents matrix of 3*3, 4 represents matrix of 4*2, 5 represents matrix of

4*3, and 6 represents matrix of 4*4. With the increasing dimensions of the sample matrix, the two kinds of algorithm have greatly improved recognition rate. Under the same dimension, RDW - LBP algorithm recognition rate is superior to the recognition rate of LBP algorithm. After 3*3 matrix, the recognition rate is not obviously increased with the increase of dimensionality, but the increase of dimensionality decreases the speed of operation which will not be able to improve the efficiency. In the experiment, 3*3 matrix is selected as the sample matrix.
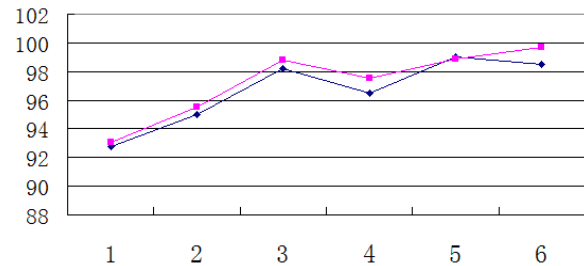

Figure 5.   Recognition rates with different matrix by the LBP and RDW-LBP in the AR face
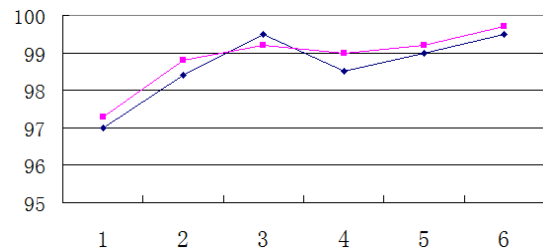

Figure 6.   Recognition rates with different matrix by the LBP and RDW-LBP in the ORL face
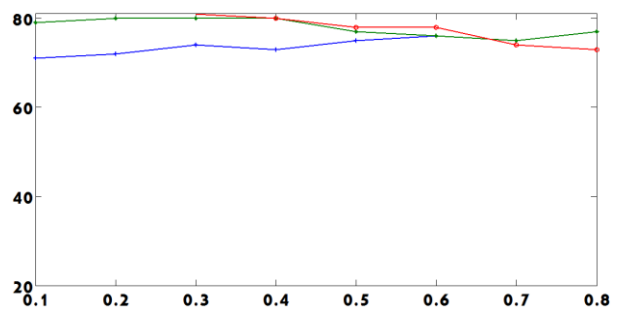
Figure 7.   Recognition rate with noise in the Yale face recognition

The recognition rate with noise in the yale face database is shown in Fig. 7 in which the red line represents the PCA algorithm, the green line represents LBP algorithm, and the blue line represents the RDW-LBP algorithm. Horizontal axis represents variance of Gaussian noise and vertical axis represents recognition rate. It clearly shows that the recognition rates with noise by PCA and LBP decrease with the increase of noise, but the recognition rate by RDW-LBP increases. With the Feret face database, the recognition rates with the three different algorithms have the same situation as is shown in Fig. 8. The experiments show that the RDW-LBP algorithm has good robustness.

TABLE III.   RECOGNITION RATE OF DIFFERENT ALGORITHMS IN THE AR FACE RECOGNITION

	PCA		LBP		RDW-LBP	
	With noise	Without noise	With noise	Without noise	With noise	Without noise
Recognition rate	93.2%	94.2%	95.5 %	97.1%	97.3%	98.4%

TABLE IV.   RECOGNITION RATE OF DIFFERENT ALGORITHMS IN THE ORL FACE RECOGNITION

	PCA		LBP		RDW-LBP	
	With noise	Without noise	With noise	Without noise	With noise	Without noise
Recognition rate	92.7%	93.5%	95.2%	96.3%	98.2%	98.7%

V.   CONCLUSIONS

In order to increase the performance of face recognition algorithm, this paper has proposed an improved algorithm.

It has been a major research goal to improve the recognition rate and robustness of face recognition algorithm. Aiming at this goal, this paper first analyzes the influence of the directional details in face image and proposes the HCPP. Combining with HCPP, this paper puts forward a directional weighted based multi-scale regional type specification of binary local RDW-LBP texture description operator for robust face recognition algorithm. The Chi-Square is used to classify extracted characteristic vector. Experimental results indicate that HCPP is reasonable and correct, relative to the LBP operator. RDW-LBP operator proposed in this paper does not increase the calculation complexity, but effectively improves the recognition rate.
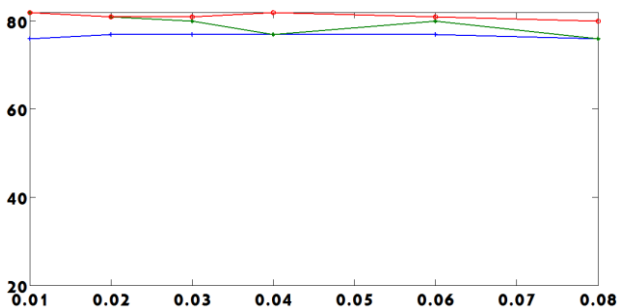

Figure 8.   Recognition rate with noise in the Feret face recognition

REFERENCES

[1]   Rabia Jafri, Hamid R. Arabnia. "A Survey of Face Recognition Techniques," *Journal of Information Processing Systems,* 2009, 5(2) pp. 41-68.

[2]   OJALA T, PIETIKAINEN M, HARWOOD D. "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition,* 1996, 29(1) pp. 51-59.

[3]   T. Ojala, M. Pietikainen, and M. Maenpaa. "Multiresolution gray-scale and rotationinvariant texture classification width local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2002, 24(7) pp. 971-987.

[4]   Timo A, Abdenour H, Matti P. "Face recognition with local binary patterns," *In proceedings of the 8th European Conference on Computer Vision(ECCV'2004), Prague: Czech Republic,* 2004, 469-481.

[5]   Timo A, Abdenour Hadid, Matti Pietik äinen. "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2006, 28(12) pp. 2037-2041.

[6]   Shengcai Liao, Xiangxin Zhu, Zhen Lei, Lun Zhang, Stan Z. Li. "Learning multi-scale block local binary patterns for face recognition," *In Proceedings of IAPR/IEEE International Conference on Biometrics (ICB'2007), Seoul: Korea,* 2007, 828-837.

[7]   Xiaoyang Tan and Bill Triggs. "Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions," *In Proceedings of the 2007 IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'2006), Rio de Janeiro, Brazil,* 2006, 168-182.

[8]   Tan X and Triggs B.. "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing,* 2010, 19(6) pp. 1635-1650.

[9]   Rafael C. Gonzalez, Richard E. Woods. "Digital Image Processing (Second Edition)," *USA: Prentice Hall,* 2002.

[10] Wang, H., Li, S., Wang, Y. "Face recognition under varying lighting conditions using self-quotient image," *In Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (AFGR'2004). Seoul: Korea,* 2004, 819-824.

[11] Jobson D, Rahman Z, Woodell G. "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image 128Processing,* 1997, 6(7) pp. 965-976.

[12] Weilong Chen, Meng Joo Er. "Illumination compensation and normalization for robust face recognition using discrete cosine transform," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* 2006, 36(2) pp. 458-466.

[13] Xiaoguang Lu and Anil K. Jain, "Automatic Feature Extraction for Multiview 3D Face Recognition," *In Proceedings of the 7th IEEE International Conference on Automatic Face and Gesture Recognition (FG'2006). Southsampton: UK,* 2006, 585-590.

[14] Khaki K, Stonham TJ. 2012. "Face recognition with weightless neural networks using the MIT database," *3rd International Conference on Autonomous and Intelligent Systems,* AIS 2012, June 25, 2012 - June 27, 2012. Springer Verlag; p. 228–33.

[15] Zhiming Qian, Dan Xu. "Automatic eye detection using intensity filtering and K-means clustering," *Pattern Recognition Letters,* 2010, 31(12) pp. 1633-1640.

[16] Zhimin Cao, Qi Yin, Xiaoou Tang, Jian Sun. "Face recognition with Learning-based Descriptor," *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR'2010). San Francisco: USA,* 2011, 2707-2714.

[17] Yuntao Wu, Bingzhi Dong, Shengji Xia, "Neural Network Optimized by an Improved PSO and Its Application in the Forecast of Nanofiltration Membrane Flux," *Advances in Information Sciences and Service Sciences,* 2012, Vol. 4, No. 22, pp. 641- 648.

[18] ángel Serrano, Isaac Martín de Diego, Cristina Conde, et al. "Recent advances in facebiometrics with Gabor wavelets: A review," *Pattern Recognition Letters,* 2010, 31(5) pp. 372-381

[19] Tan X and Triggs B.. "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing,* 2011, 19(6) pp. 1635-1650.

[20] O. Déniz, G. Bueno, J. Salido, F. De la Torre. "Face recognition using Histograms of Oriented Gradients," *Pattern Recognition Letters,* 2011, 32(12) pp. 1598-1603.

[21] SHU Chang, DING Xiaoqing, FANG Chi. "Histogram of the Oriented Gradient for Face Recognition," *Tsinghua Science and Technology,* 2011, 16(2) pp. 216-224.

[22] N. Sun, H. -X. Wang, Z. -H. Ji, C. -R. Zou, and L. Zhao. "An efficient algorithm for Kernel two-dimensional principal component analysis," *Neural Computing & Applications,* 2008, 17(1) pp. 59-64.

[23] WANG Lin, LI Yong-ping, WANG Cheng-bo, et al. "2D Gaborface representation method for face recognition with ensemble and multichannel model," *Image and Vision computing,* 2008, 26(6) pp. 820-828.

Color Image Segmentation Method Based on Improved Spectral Clustering Algorithm

Dong Qin

Yancheng Institute of Technology, Yan Cheng, China

Email: dongqin011@163.com

Abstract—**Contraposing to the features of image data with high sparsity of and the problems on determination of clustering numbers, we try to put forward an color image segmentation algorithm, combined with semi-supervised machine learning technology and spectral graph theory. By the research of related theories and methods of spectral clustering algorithms, we introduce information entropy conception to design a method which can automatically optimize the scale parameter value. So it avoids the unstability in clustering result of the scale parameter input manually. In addition, we try to excavate available priori information existing in large number of non-generic data and apply semi-supervised algorithm to improve the clustering performance for rare class. We also use added tag data to compute similar matrix and perform clustering through FKCM algorithms. By the simulation of standard dataset and image segmentation, the experiments demonstrate our algorithm has overcome the defects of traditional spectral clustering methods, which are sensitive to outliers and easy to fall into local optimum, and also poor in the convergence rate.**

Index Terms—**Spectral Clustering; Segmentation; Intimate Matrix; Scale Parameter; KFCM**

I. INTRODUCTION

Color image segmentation based on data clustering is to take the whole image as an integral data. The predetermined defined distance measure method is used to calculate the similarity relation of pixel point to determine different regions of image. During past dozens of years, people apply many different clustering algorithms into image segmentation such as means algorithm [1, 2]. Since K means algorithm only fits data in spherical space and many data structures in reality are not this shape, segmentation results which are caused by these methods cannot be satisfied. Recently, spectral clustering based on spectral graph theory [3, 4, 5] has attracted people's focus of research. Compared to traditional algorithm, it has obvious advantages. This algorithm can operate in sample space with random shape and converge in global optimization and this characteristic extensively adapts to data. In addition, by means of spectral clustering, Gauss kernel needs to be applied to calculate similarity between random two points to form similarity matrix for clustering. When the scale of data set is large, computational complexity and memory capacitance of eigenvector decomposition which is performed by L matrix obtained from similarity matrix

will become bottleneck. Meanwhile, parameter ξ in Gauss kernel has the largest influence on similarity between two points. However, there is not any effective method in current theory to automatically select the most appropriate value [6]. The value can only be set according to human experience. If data set type is different, ζ value is also different so it causes instable clustering results. On the basis of above analysis, there are two points which can be generalized from problems in this method: (1) Clustering is very sensitive to parameter; (2) Large time complexity and space complexity.

Towards above problems, scholars constantly study and effectively combine ensemble learning methods and spectral clustering [7, 8] to improve the robustness of clustering method and to avoid parameter selection troubles. Meanwhile, through many studies, it has been proved that the performance of clustering ensemble is superior to single clustering algorithm when it performs data-clustering with random shape and scale [9]. In 2003, Dudoit and J.Fridlyand [10] applied BOOTSTRAP sampling technology of data to generate different subsets of data and used the same algorithm to cluster on these different subsets so as to form a group of clustering. In 2005, Tang Wei and Zhou Zhihua [11] applied generating method of selective clustering member. They used BOOTSTRAP sampling method to obtain different data subset and cluster in this data subset with clustering algorithm to get different clustering members. Muna and Domeniconi [12] used different parameters of Locally Adaptive Clustering algorithm to generate diversified clustering members. Through certain selective strategy, Fern and Lin [13] selected some of clustering results to ensemble. In 2009, Hore, etc divided large dataset into non-overlapping data subset and clustered each data subset. The obtained diversified data is divided and BIPARTITE and METIS methods are used to integrate. Jia [14] applied the nearest neighbor method to select suitable clustering member to obtain integration result. Through above introductions, we can see that scholars adopted different methods to study based on image segmentation and obtain satisfied results to some extent. The implementation clustering method to study image segmentation attaches scholars' more and more attentions.

Therefore, this paper uses spectral clustering method to deeply study the image segmentation and clustering integration technology to further improve the performance of image segmentation. We attempt to

excavate available prior information in amounts of non-category data and apply semi-supervised algorithm to solve and excavate rare class problems. As image segmentation method is too sensitive to parameter selection, ensemble learning method with robustness and strong generalization ability can be applied. A self-adaptive semi-supervised fuzzy spectral clustering algorithm is proposed. This algorithm does not need to determine clustering category and it computes the value of scale parameter σ according to the distance between two points. Through the analysis on the spectral clustering algorithm and the problems of super-boundary integration method, K-MEANS algorithm in spectral clustering is replaced by FKCM algorithm. The fuzzy cluster validity function is introduced in algorithm to optimize cluster number. Membership matrix formed by FCM will extend "0" and "1" in super boundary fusion method to range [0,1] and cluster new super boundary so as to obtain final image segmentation result. Finally, by means of simulation experiment of data sets and image segmentation, the result shows that clustering effect of this algorithm is perfect.

II. PRELIMINARY WORKS

A. Image Segmentation Models

Image segmentation technology has become a research spot in the field of image processing to be studied by most scholars recently. Thousands of image segmentation algorithms have been proposed so far and hundreds of research results will be published every year. Since most image segmentation algorithms are put forward for special problems, there are big limitations and pertinence [15]. Related researchers have proposed an accepted definition for image segmentation based on the set theory [16]:

The images to be segmented are looked as a set denoted by R. The process of segmentation is equivalent to segment R into several sub-sets $R_1, R_2, ..., R_n$ by different algorithms. The sub-sets should satisfy the following conditions:

(a) $\bigcup_{i=1}^{n} R_i = R$

(b) $R_i \cap R_j \neq \Phi$

(c) $P(R_i) = TRUE, i = 1, 2, ..., n$

(d) $P(R_i \cup R_j) = FALSE, i \neq j$

(e) R_i is connected area, $i = 1, 2, ..., n$

In the above points, (a) means the segmented images set can be used to revert to original images; (b) means there are not overlapped parts among the segmented sub-sets; (c) means each of the segmented sub-sets should be non-empty; (d) means different segmented sub-sets are independent each other in character or feature; (e) means any of segmented sub-sets is connected.

B. Implementation of Spectral Clustering Algorithm

The idea of spectral clustering algorithm originates from spectra division theory and it takes clustering as a multi-channeled division problem of undirected graph.

Original sample data is mapped to k-dimension space. Then, clustering algorithms such as traditional k-means, kernel c-means, etc are applied to cluster in k-dimension space.

We assume a dataset X, $X = \{x_1, x_2, ..., x_k\}$ is considered as an undirected weighted graph $G(V, E, A)$. V is vertex set, E is side set, A is intimacy matrix set which denotes the side weight set to connect two vertex. We choose Gauss kernel function to define the intimacy matrix

$$\bar{k}_{ij} = \exp(-\frac{\| x_i - x_j \|^2}{\sigma^2}), i \neq j \qquad (1)$$

$\bar{k}_{ij}$ is kernel of Gauss, σ is kernel radius which controls the attenuation speed of $\bar{k}_{ij}$. According to spectrum theory the value of σ is given and the line vector of $\bar{k}_{ij}$ will surround them for clustering which distribute on the hypersphere of k-dimension space. D is diagonal matrix and

$$D_{ij} = \sum_{j=1}^{n} \bar{k}_{ij} \qquad (2)$$

$$k_{ij} = \frac{\bar{k}_{ij}}{\sqrt{D_i D_i}} \qquad (3)$$

Calculate corresponding eigenvectors of the first k maximum eigenvalues in matrix K to establish matrix $Y[n \times k]$. Each line of matrix Y is unitized to get matrix V

$$V_{ij} = Y_{ij} / (\sum_j Y_{ij}^2)^{\frac{1}{2}} \qquad (4)$$

All the line vectors in V are considered as points of R^n-dimension space. These data points are clustered with k-means or other clustering methods. If and only if the i_{th} behavior is clustered as j, the original data point x_i is the j_{th} class.

Spectral clustering often constructs weighted graph using the image pixels as convexes when segmenting images. The weight of side reflects the similarity among the pixels. By analysis on the eigenvectors and eigenvalues of weighted matrix whose elements are the side weight of graph, the vertexes can be clustered to segment the image. For a piece of image, the principle rules for segmentation are to make maximum internal similarity of k sub-images and minimum similarity among these sub-images. So we can calculate the value of RAssoc [17] to determine the performance of segmentation.

$$RAssoc(G) = \max_{v_1, ..., v_k} \sum_{c=1}^{k} \frac{link(V_c, V_c)}{|V_c|} \qquad (5)$$

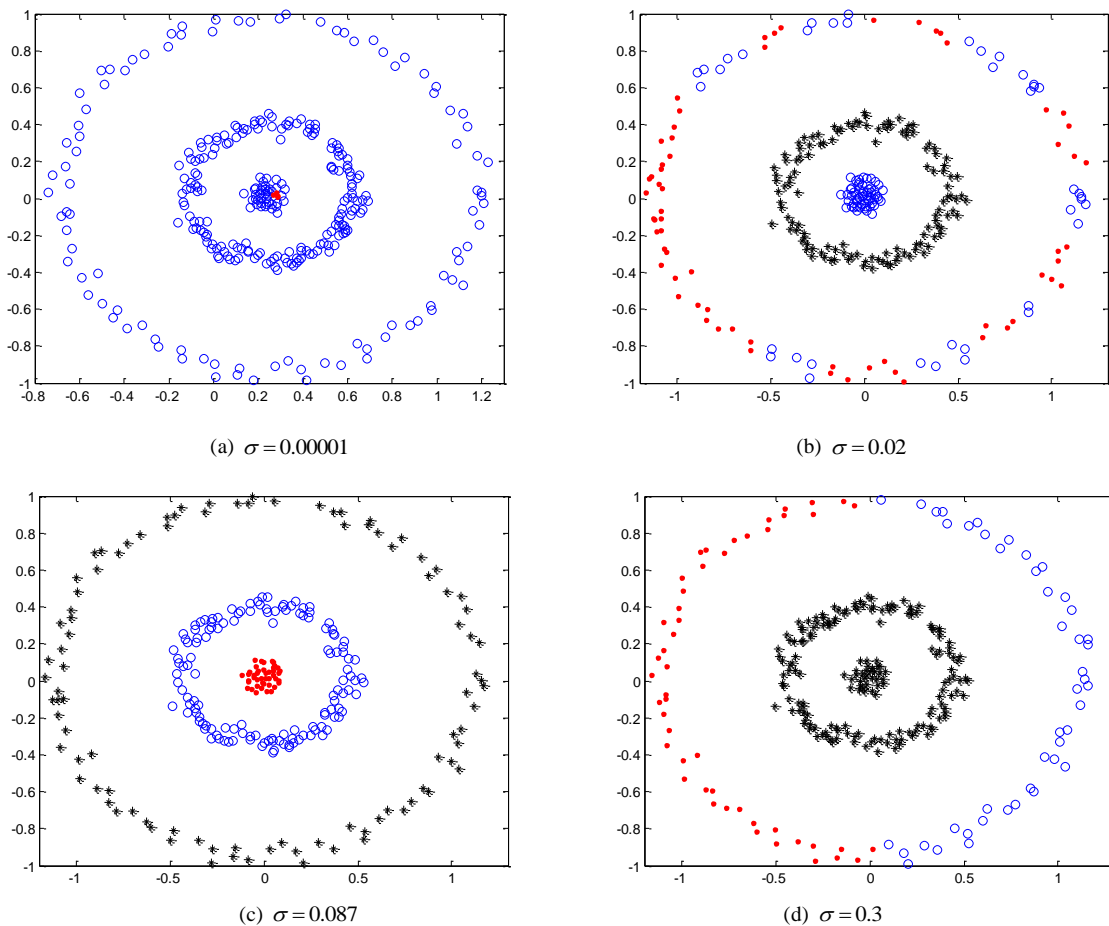$$link(A, B) = \sum_{i \in A, j \in B} W_{ij}$$

(a) $\sigma = 0.00001$

(b) $\sigma = 0.02$

(c) $\sigma = 0.087$

(d) $\sigma = 0.3$

Figure 1.   Influence on clustering results of selection for σ

$\deg ress(A) = link(A, V)$

W_{ij} denotes the connected weight among the pixels.

III.   ADAPTIVE FUZZY SEMI-SUPERVISED SPECTRAL CLUSTERING ALGORITHM

A. Optimization of Scale Parameter

It is san important procedure to establish intimate matrix in spectral clustering algorithm. Usually Gauss kernel function is used to define the intimate matrix. For the scaling parameters the given kernel radius will directly affect the clustering results. The scaling parameter may be looked as a measuring standard to judge whether two data points are the same, as is considered in data processing. While there are not corresponding provided standard to process the data belonging to different fields. So the determination of kernel radius will influence the final results by scaling parameter σ, as shown in the following experiments.

It can be seen that the clustering results are different with different σ, so the choice of σ can not be ignored. The relation between c and clusters number is: when σ has smaller value, which means each data object has smaller effect on around areas. w_{ij} appears to the sum of spike-like functions who takes n data points as the center. The clustering result will generate many clusters. The

most extreme case is that each data point aggregates into a cluster respectively, which is obviously pointless. On the contrary, when σ takes a larger value, each data object has bigger effect on around areas. w_{ij} appears to the sum of function composed by n data points, changing slowly with large width. In the most extreme case, all the points are aggregated into one cluster. Therefore, to acquire clustering results as accurate as possible, the decision of σ should embody the distributed features of original data.

Zelnik-Manor proposed an adaptive SC algorithm in [18] to solve the problem of σ. The optimal determination of σ in spectral clustering can be performed as the searching method directly on original data. The cross-validation may be used to avoid the problem caused by single choice of scaling parameter σ. In this paper we will replace the calculation of σ with local σ_i of each data point. So the distance between x_i and x_j is $d(x_i, x_j)/\sigma_i$ and the contrary from x_j to x_i is $d(x_j, x_i)/\sigma_i$. Then the distance between two points is defined as:

$$d^2(x_i, x_j) = \frac{d(x_i, x_j)d(x_j, x_i)}{\sigma_i \sigma_j} \qquad (6)$$

The intimate matrix is:

$$k_{ij} = \exp[-\frac{1}{\sigma_i \sigma_j} d^2(x_i, x_j)] \tag{7}$$

Its diagonal matrix and Laplacian matrix are:

$$D_{ii} = \sum_{j=1}^{n} k_{ij} \tag{8}$$

$$K_{ij} = \frac{k_{ij}}{\sqrt{D_i D_j}} \tag{9}$$

We use a special scaling parameter which permits them to adaptively regulate the distance between the points according to the ne local statistics of neighborhood. The choice of σ_i can be determined by the research on local statistics of data point x_i. So σ_i is calculated as:

$$\sigma_i = d(x_i, x_m) \tag{10}$$

x_m is the m_{th} neighbor point of x_i and $d(x_i, x_m)$ is the distance between x_i and x_m.

B. Determination of Validity Index

The clustering number c is closely related to the clustering quality: larger value will make the clustering results complicated and hard to explain and analyze; smaller value will cause information loss and misleading to the final decision. The validity index can be used to determine reasonable clustering numbers, to ensure more effective clustering results of the algorithm. To acquire better fuzzy kernel c-means results, we choose the kemelized AmineM.Bensaid validity index. It is defined as:

$$V_{KBsaid}(U, V; c) = \sum_{i=1}^{c} [\frac{\sum_{j=1}^{n} u_{ij}(1 - K(v_i, x_j))}{n_j \sum_{i=1}^{c}(1 - K(v_i, v_j))}] \tag{11}$$

$n_j = \sum_{j=1}^{n} u_{ij}$, $\sum_{j=1}^{n} u_{ij}(1 - K(v_i, x_j))$ are used to measure the compactness inside the class. The smaller this value is, the more compact the element in the class is; $n_j \sum_{i=1}^{c}(1 - K(v_i, v_j))$ is to measure the separation among the classes.

C. KFCM Algorithm

Given limited samples set $x_k \in R^n$, kernel function Φ is used to map the samples set on high-dimension space H, that is, $\Phi: R^n \to H$. The topology structure of samples in the original space keeps unchanged and $\Phi(x_k)$ denotes the k_{th} class center of feature space H. Then the target function of FKCM clustering is:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{\alpha} \| \Phi(x_i) - \Phi(v_i) \|^2$$
$$= \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{\alpha} [K(x_j, x_j) - 2K(x_j, v_j) + K(v_i, v_i)] \tag{12}$$

v_i denotes the class center of the i_{th} class; $\alpha > 0$ denotes the weight index; $\Phi(v_i)$ denotes the image of this center in corresponding kernel space and it is described as:

$$\Phi(v_i) = \frac{\sum_{j=1}^{n} u_{ij}^{\alpha} \Phi(x_j)}{\sum_{j=1}^{n} u_{ij}^{\alpha}} \tag{13}$$

Satisfying following constricts:

$$\begin{cases} 0 \le \mu \le 1, i = 1, 2, ..., n, j = 1, 2, ..., k \\ \sum_{j=1}^{k} \mu_{ij} = 1 \\ 0 < \sum_{i=1}^{n} \mu_{ij} < n \end{cases}$$

In the feature space H, the membership function is:

$$\mu_{ij} = \frac{1}{\sum_{g=1}^{c} (\frac{Q_{ij}}{Q_{gi}})^{\frac{1}{\alpha-1}}} \tag{14}$$

$Q_{ij} = K(x_j, x_j) - 2K(x_j, v_j) + K(v_i, v_i)$ is the Euclidean distance from the j_{th} sample to the center of the i_{th} class.

$$K(x_k, v_i) = \Phi(x_k)\Phi(v_i) = \sum_{j=1}^{n} u_{ij}^{\alpha} K(x_j, x_k) / \sum_{j=1}^{n} u_{ij}^{\alpha} \tag{15}$$

$$K(v_i, v_i) = \Phi(v_i)\Phi(v_i) = \sum_{j=1}^{n} \sum_{l=1}^{n} u_{ij}^{\alpha} u_{il}^{\alpha} K(x_j, x_l) / [\sum_{j=1}^{n} u_{ij}^{\alpha}]^2 \tag{16}$$

So we get

$$Q_{ik} = K_{kk} - (2/N)\sum_{j=1}^{n} \mu_{ij} K_{kj} + (1/N_i^2)\sum_{j=1}^{n} \sum_{l=1}^{n} \mu_{ij} \mu_{il} K_{jl} \tag{17}$$

Bezdek introduces the weighted index m, called smoothing factor either, to control the sharing degree among the fuzzy clustering. Thus, to acquire the fuzzy clustering we must choose suitable m. Since the selection of optimal m lacks theoretic instruction, Bezdek and Hathway study the conclusion about sample number n which is related to the value of m, from the convergence of algorithm. They advice $m > n/(n-2)$. In Pal's research, in the experiments of clustering effectiveness, the conclusion is that the optimal range of m should be [1.5, 2.5] and $m = 2$ is generally adopted.

The detail of KFCM algorithm is described as the follows:

Step 1: Given the number of clustering class $c(2 \leq c \leq k)$. k is the number of samples. Assuming the iteration ending threshold $\varepsilon = 0.01$ and the algorithm iteration calculator $t = 1$. Choose the kernel function K and its parameters;

Step 2: Initializing the membership matrix;

Step 3: Computing the distance from samples to the clustering center according to equation 17;

Step 4: Recalculating the membership degree of each sample according to equation 14;

Step 5: If $\max_{i,j} | \mu_{ij}^t - \mu_{ij}^{t-1} | > \varepsilon$, go to step 3;

Step 6: If $V_{KBsald}^t > V_{KBsald}^{t-1}$, the validity index is the optimal clustering number c; otherwise $c = c + 1$ and go to step3.

D. Adaptive Fuzzy Spectral Clustering Algorithm

The adaptive semi-supervised fuzzy spectral clustering algorithm proposed by our paper is using the data information of tag to adjust the distance matrix formed by the distance among the points. It makes the points distribution inside the clustering as close as possible and the clustering as separated as possible each other. When adjusted matrix adopts the spectral clustering algorithm for clustering, we should make the distance between the K_{th} and $(K+1)_{th}$ eigenvalue of matrix K as large as possible, to acquire better clustering results.

In the semi-supervised clustering process, the tag data provided by users have two kinds commonly [19]: Tag data with class; pair of point limit (marking whether two points are from the same class). The tag data of the latter form is some kind of weaker tat data, but the former is more reasonable for tag data with classes. We can turn it into equivalent pair point limit, but not vice versa. In our algorithm, the data form of tag is pair point limit. There are

There are two sets $M = \{(x_i, x_j)\}$ and $C = \{(x_i, x_j)\}$ (x_i, x_j of C belong to different cluster), $Y = [y_1^T, y_2^T, ..., y_k^T]$. Matrix V is acquired by the normalization of matrix Y.

$$\cos \theta_{ij} = \frac{y_i^T y_j}{\| y_i \| \| y_j \|} = \begin{cases} 1, (x_i, x_j \in M) \\ 0, (x_i, x_j \in C) \end{cases} \quad (18)$$

y_i is the i_{th} line vector of Y. According to above theories, we may calculate the quality S_k with the clustering results.

$$S_k = \sum_{c=1}^{k} \frac{1}{N_c} \sum_{i,j \in M_c} \cos \theta_{ij} \quad (19)$$

M_c is the data point set in class c and N_c is the number of data points. When the values of S_k is bigger, the similarity of data point in each class if higher. So the value of k makes maximum S_k is the optimal clustering number.

For spectral clustering algorithm, we only need the distance function satisfy the symmetry. When $x_i, x_j < M$, adjust $D_i = D_j = 0$. To ensure the distance among the points is a reasonable norm, we adopt the shortest path algorithm to modify the distance among all the related points in matrix; when $x_i, x_j < C$, adjust $D_i = D_j = \infty$. We adjust the distance matrix according to the tag data and use the adjusted matrix for spectral clustering. Since original distance matrix is the Euclidean distance in space, it satisfies the triangle inequality. So any two points which need to adjust the distance between 2 points must be the point via set I ($I = \{i : \exists j \neq i, (i, j) \in M\}$). Then the algorithm of acquired the shortest path can be obtained rapidly.

During the process of spectral clustering, when matrix Y is performed FCM clustering, we can determine the original point as the following method: Select a point as the center point of the first cluster randomly. Choose the points which are orthogonal to all the center points possibly as the center point for the next clustering, until the k_{th} original center points is found. The procedures of improved algorithm are show as follows:

Step 1: Calculate the scaling parameter σ_i of each data point $x_i \in X$;

Step 2: For any $x_j \in X$, calculate the euclidean distance of two points D_{ij};

Step 3: Add pair limit information and adjust the distance matrix D according to tag data;

$$\begin{cases} D_{ij}, D_{ji} = 0, \ if (x_i, x_j) \in M \\ D_{ij}, D_{ji} = \infty, if (x_i, x_j) \in C \end{cases}$$

Step 4: Establish similarity matrix $\hat{k}_{ij}$ based on adjusted D;

Step 5: For $k = 2$ to $k_{\max}$

a) Establish matrix K and calculate the corresponding eigenvectors to the previous k_{th} maximum eigenvalues of $\hat{k}_{ij}$ { $v_1, v_2, ..., v_k$ };

b) Establish matrix Y and normalize K to Y;

c) Calculate $\cos \theta_{ij}$ and S_k;

Step 6: Fine suitable k_{opt} to make the maximum S_k;

Step 7: Set each line in Y as some point in R^k and clustering it into k_{opt} class with FCM;

Step 8: If the i_{th} line of Y belongs to the j_{th} class, then divide original point x_i into the j_{th} class.

IV. EXPERIMENTAL RESULTS

A. Standard Datasets

Experimental data are mainly from Iris data set and Wine data set in UCI. Iris data set is the most famous UCI real data set to inspect clustering algorithm and validity index performance. Corresponding to Iris data set, Wine is another famous data set in UCI database. This

dataset is made up of 178 samples in 3 types and each sample has 13 attributes. Towards high dimensional data set, the determined optimal clustering number in the whole range of $m < [1.5, 2.5]$ is stable. Restricted numbers are from 0 to 299 in all experiments. Each given limited number is performed 20 experiments. Since selected restriction of output average results is different, clustering algorithm attribute will have large influence. Following methods are adopted to produce limitations for experimental equality: the same limited number will produce 100 groups of different limitations. For this reason, different random generator of seeds initialization is performed. Seed number increases from 1 to 110 so different limited sets under the same limited number can be obtained.
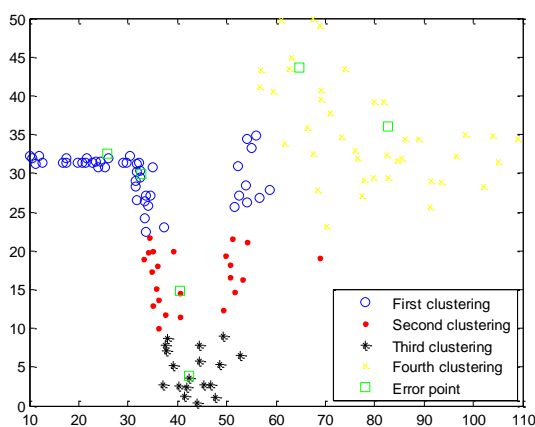

Figure 2. Clustering result of improved algorithm

This experiment applies three clustering methods. The first clustering method adopts traditional k-means clustering and makes 10 times of iterations. Each obtained clustering result is respectively compared to classification result of data set in UCJ database and average value of accuracy sum from these 10 times clustering results are taken as its final clustering accuracy. The second clustering method adopts NJW spectral clustering algorithm. 10 a values are randomly selected from 0 to 10. Each obtained clustering result is individually compared to classification result of data set in UCI database and the average value of accuracy sum from these 10 times clustering results is taken as its final clustering accuracy. The third clustering method adopts EBSC method proposed in this paper without manually setting scale parameter 6. The obtained clustering result is compared to classification result of data set in UCI database to calculate its clustering accuracy.

TABLE I.　　CLUSTERING ACCURACY OF 3 ALGORITHMS IN UCI DATASET

Dataset	k-means	NJW	ASC
Iris	57.33	59.33	63.33
Wine	57.30	61.24	62.66
Parkinsons	73.85	74.87	76.30
Musk	51.68	57.98	60.05
Spectf	62.55	67.19	68.65
Breast-cancer	85.41	76.09	79.20
Average rate	64.69	66.12	68.37

The experimental results are shown as table 1. We can see that our algorithm is superior to previous two algorithms on clustering average accuracy. The improved algorithm operates normally and the clustering result does not rely on careful selection of user parameter, which has effective robustness.

B. Standard Image Segmentation

In order to testify algorithm effectiveness, we will carry out experiment on real image and compare it with algorithm result in reference [9]. Figure 5 presents segmentation results comparison of these two algorithms. Figure 5(a) is source image with size of 340×500, figure 5(b) is the segmentation result of reference [9] and figure 5(c) is the algorithm segmentation result in this paper.

From the experimental results we know, on image segmentation and under the condition of the same category number, the time consumption of algorithm in reference [9] and our algorithm is very large but the result of the improved algorithm is obviously better. Limited artificial information is applied during experiment. That is, the limited points are clicked on image to mark relationship between corresponding regions, so as to satisfy the given limited accurate image segmentation result. We can discover from experiment that consumed time and pixel value are rising in linearity. The selected image pixel value in this paper is approximately 180000. When this image size is smaller than 100×150, the consumed time to calculate intimacy matrix is smaller than 15. The consumed time of selected image to calculate intimacy matrix is 10.3281s in this paper. Our algorithm also has disadvantages during experiment process. Such as large pixel figures, when the eigenvector and eigenvalue of weighted matrix are calculated, their operations are a bit slow in comparison with smaller images. We can see from this result that each block is mutually independent region which can express comprehensive information and conveniently extract high-level information

Table 2 is the comparison of these two algorithms in consumed time and RAssoc value. From this table, this paper introduced weighted fuzziness to reduce consumed time. In addition, RAssoc value refers to interior approximate degree of sub-image. The larger the value, the higher the approximate degree inside sub-image, the better the segmentation result. Obviously, segmentation results of the improved algorithm are superior to those in reference [9].

TABLE II.　　COMPARISON OF THE ALGORITHM PARAMETERS

Algorithm	Consumed time	Rassoc
Adaptive spectral clustering	8.339	15.997
Algorithm in reference [9]	117.898	15.825

To testify the performance of clustering integration method on large size of image segmentation, the available images in experiment are shown as figure 6. Image (a) and (b) are images taken on their own, the image size is 1024×768 (modified as 480×320 in experiment). Others are from Berkeley image segmentation database and image size is 480×320 (source image is 481×321). Ncut

(a) Original image   (b) Rewritten image   (c) Improved segmentation
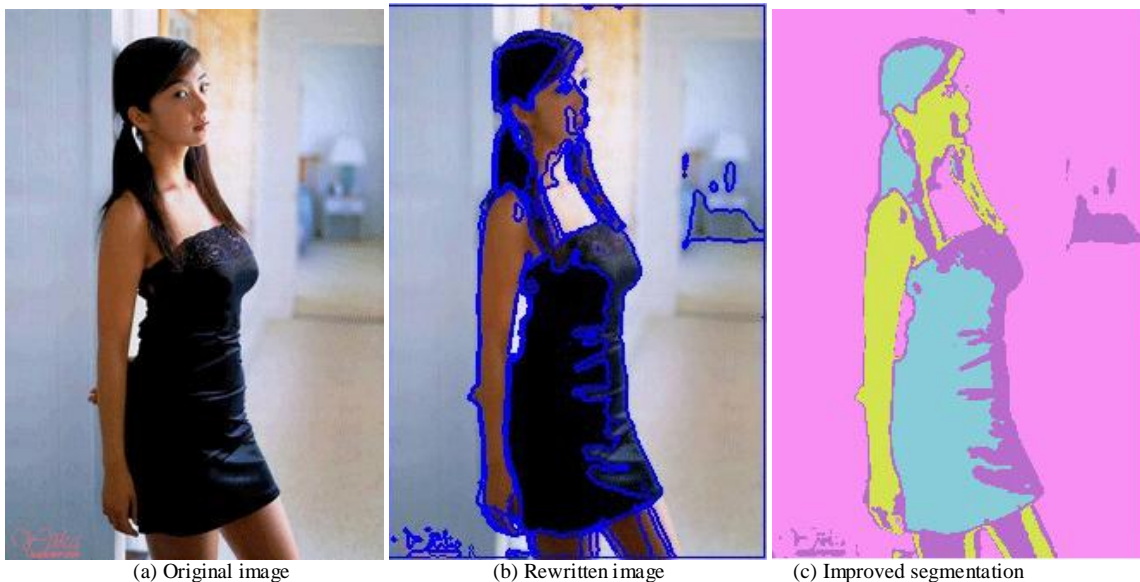
Figure 3.　Segmentation results comparison
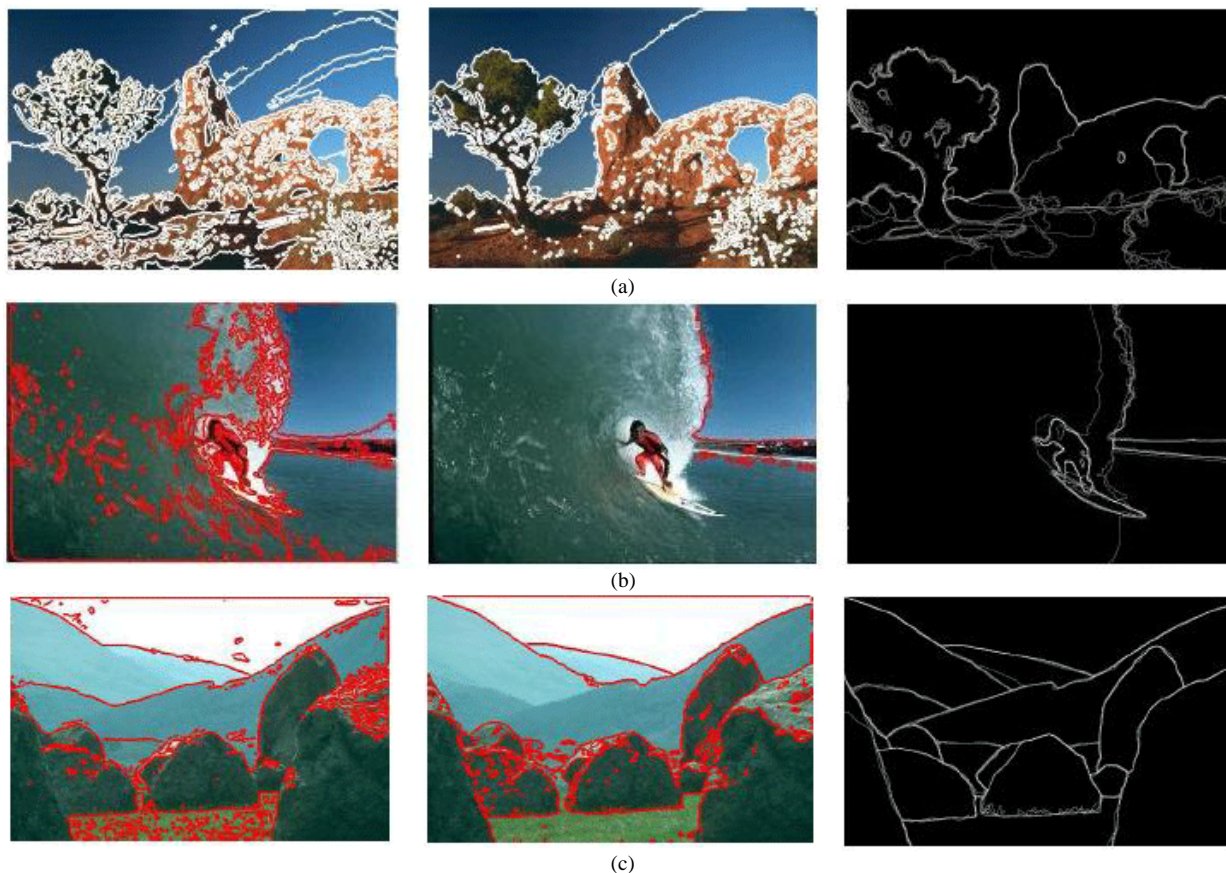

(a)

(b)

(c)

Figure 4.　The results of three segmenting algorithms

method in experiment and method in previous experiment are used for reference. They show the experimental results. Image (a), (b), (c) respectively refer to integrated segmentation result of images, Ncut segmentation result, Ncut segmentation result and segmentation result based on blocking technology segmentation algorithm. In this experiment, $m = 2$, ζ is randomly selected in $[0.01\ 0.09]$ and the clustering member number is 7.

We can see that the method in this paper is not inferior to Ncut method and it is superior to Ncut in some details. For example, the detail of tree in 3a, the bear's outline detail in 4a and the architectural outline in 5a are all shown clearly. The most important point is that the processing speed of our method is obviously higher than Ncut. When image scale is large, Ncut time consumption is several times as much as this paper and this paper has ability to process large scaled image.

V. CONCLUSION AND FUTURE WORK

The main work of our research is semi-supervised spectral clustering and its application in image segmentation. This paper mainly studies spectral clustering algorithm and application of fuzzy C-means algorithm in image segmentation technology. Based on characteristics and how to determine clustering quantity, this paper attempts to explore a combination algorithm between semi-supervised machine learning technique and spectral graph theory. For the improved self-adaptive semi-supervised fuzzy spectral clustering algorithm, its scale parameter optimizes through self-adaptive technology based on information entropy. Then through adding taged data, the taged data are used to regulate distance matrix, calculate similar matrix, discover suitable k and finally cluster through fuzzy c-means. By means of simulation experiments on standard data sets and image segmentation, experimental results show that the improved algorithm can effectively improve clustering quality.

We also need further improvements and studies and the following points are priorities for future research:

- To provide universal principles and construct similarity function
- To utilize knowledge of image segmentation and fuzzy theory and construct fuzzy segmentation target function
- According to practical problems, technologies in many fields are integrated specifically to improve classification performance.

REFERENCES

[1] J. Shi, J. Malik, Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell,* vol. 22, no. 8, pp. 888-905, 2000.

[2] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing,* vol. 17, no. 14, pp. 395-416, 2007.

[3] X. R. Zhang, Spectral Clustering Ensemble Applied to SAR Image Segmentation. *IEEE Geoscienceand Remote Sensing Society,* vol. 46, no. 7, pp. 2126-2136, 2008.

[4] Toussi Soheila Ashkezari, Yazdi Hadi Sadoghi, Feature selection in spectral clustering, *International Journal of Signal Processing, Image Processing and Pattern Recognition,* vol. 4, no. 3, pp. 179-194, 2011.

[5] Liu Hanqiang, Jiao Licheng, Zhao Feng, Local manifold spectral clustering with FCM data condensation, *Proceedings of the International Society for Optical Engineering,* pp. 1009-1013, China, 2009.

[6] Xiaofei Li, Xuanjing Shen, Haipeng Chen, ElGamal Digital Signature Algorithm of Adding a Random Number, *Journal of Networks,* vol. 6, no. 5, pp. 774-782, 2011.

[7] Zhang Quan, Hu Yulan, Image Segmentation Algorithm Based on Spectral Clustering Algorithm, *Journal of Shenyang Ligong University,* vol. 31. no. 6, pp. 87-91.

[8] Meng Qingxia, Wen Xianbin, Xu Haixia, Improved synthetic aperture radar image spectral clustering algorithm, *Acta Optica Sinica,* vol. 32, no. 1, 2012.

[9] YIN Fang, CHEN Deyun, WU Rui, Improved method of image segmentation using spectral clustering, *Computer Engineering and Applications,* vol. 47, no. 21, pp. 185-187, 2011.

[10] J. Fridlyand, Bagging to improve the accuracy of a clustering procedure, *Bioinformatics,* vol. 19, no. 9, pp. 1090-1099, 2003.

[11] Tang W, Zhou ZH, Bagging-Based selective clusterer ensemble, *Journal of Software,* vol. 16, no. 4, pp. 496−502, 2005.

[12] M. Al-Razgan, C. Domeniconi, Weighted clustering ensembles, *Proceedings of the Sixth SIAM International Conference on Data Mining,* pp. 258-269. USA, 2006.

[13] A. Fred, A. K. Jain, Data clustering using evidence accumulation. *Proceeding of 16th Int'l Pattern Recognition,* pp. 276-280, USA, 2002.

[14] Yu Jin-hua, Wang Yuan-yuan, Shi Xin-ling, Image segmentation with two-dimension fuzzy cluster method based on spatial information, *Opto-Electronic Engineering,* vol. 34, no. 4, pp. 114-119, 2007.

[15] Mohamed Samar S., Salama Magdy M. A., Spectral clustering for TRUS images, *BioMedical Engineering Online,* vol. 15, no. 6, pp. 70-78, 2007.

[16] Yang Yuanfeng, Cui Zhiming, Wu Jian, Trajectory analysis using spectral clustering and sequene pattern mining, *Journal of Computational Information Systems,* vol. 8, no. 6, pp. 2637-2645, 2012.

[17] Jin HuiZhen, A Spectral Clustering-Based Algorithm for Gray Graph Segmentation, *Computer Systems & Applications,* vol. 4, no. 2, pp. 74-77, 2009.

[18] Bezdek J C, A physical interpretation of fuzzy ISODATA, *IEEE Transactions on Scientific Manpower Commission,* vol. 6, no. 3, pp. 2873-2890, 1976.

[19] Zhao Feng, Jiao Li-Cheng, Liu Han-Qiang, Semi-Supervised Eigenvector Selection for Spectral Clustering, *Pattern Recognition and Artificial Intelligence,* vol. 24, no. 1, pp. 48-56, 2011.

A Novel Local Features Based Salient Object Recognition Algorithm via Hybrid SVM-QPSO Model

Xin Wang, Tianzhong Zhao, and Yi Zeng*

School of Information Science & Technology, Beijing Forestry University, Beijing 100083, China
*Corresponding Author

Abstract—**As the salient objects extraction is of great importance in computer vision and multimedia information retrieval, this paper concentrates on the problem of salient object recognition using local features. Considering the rotational invariance performance of circular region is much better, we exploit a circular region to replace the rectangular region. To implement the salient object detection, the visual object classes should be constructed from training image dataset through SIFT features clustering. Furthermore, for a test image, the object class which the test image belonged to can be detected by interest points matching. Afterwards, the SIFT features clustering and local features matching process can be implemented through the proposed hybrid SVM-QPSO model. To promote the quality of parameter selection in SVM, we utilize the quantum behaved particle swarm optimization technique to select suitable SVM parameters. Finally, experiments are conducted to make performance evaluation using the MSRC dataset. Experimental results show that compared with other methods, the proposed algorithm can effectively detect salient objects in both object detecting precision and computing efficiency.**

Index Terms—**Local Features; Salient Object Recognition; SVM; Quantum Behaved Particle Swarm Optimization; SIFT; Parameters Selection**

I. INTRODUCTION

Object detection refers to a computer technology which is related to computer vision and multimedia information retrieval. High quality object detection aims to solve the problem of detecting instances of semantic objects of a certain class (such as animals, plants, or humans) in images and videos. In object detection research field, face detection and pedestrian detection are discussed mostly in recent years. Furthermore, object detection has applications in many research fields in computer vision, including multimedia information retrieval and video surveillance [1-2].

As is well known that object recognition is one of the most important problems in computer vision research field. Detecting salient object from images and videos is usually an important part in many multimedia applications such as object-based coding, object-based image/ video retrieval, image/video editing and manipulation, smart video surveillance, and human computer interaction et al [3].

As is an important research field in object detection, salient object detection [6-10] is one of the important issues in computer vision(CV), which have obtained many attentions in recent years. Particularly, visual saliency refers to the capability to detect the relevant object in an image fastly and accurately. The main works in salient object detection lie in that each pixel in the host image is given a measure of relevance score. The above process can be conducted by giving higher values to the important image regions and lower values to other image regions. On the other hand, the problem of salient object detection aims to provide a suitable solution to many complex real-time cases, including surveillance systems to track vehicle [4]. Furthermore, salient object recognition is also exploited in the application of automatic target detection, which includes 1) Finding traffic signs along the road or military vehicles in a savanna, 2) Finding salient objects in the environment as navigation landmarks in robotics.

To the best of our knowledge, the traditional studies about salient object recognition mainly focus on the global visual features of images, such as color features, shape features and texture features. However, it is an effective way to describe the salient objects using local features. Different from the traditional methods, in this paper, we focus on salient objects detecting with SIFT features, which refers to an algorithm in computer vision to detect and describe local features in images. Particularly, SVM was initially published by David Lowe in 1999.

On the other hand, classification problem is of great importance in Machine Learning, and SVM is a powerful computing tool box in it. SVM belongs to a kind of supervised learning models utilizing related learning algorithms which can analyze data or recognize patterns. Supposing there is a set of training examples, in which each one is annotated with the category name. Furthermore, the SVM training process refers to construct a model which can allocate new examples into one class or not, and then construct a non-probabilistic binary linear classifier. Intuitively, the main idea of SVM can be described as that example is represented as points in the space, and then these points are mapped. The

optimization objective of SVM is to separate the examples into different classes with a clear gap which is as wide as possible. Afterwards, new examples are then mapped into that same space, and which class the example is belonged to can be predicted depending on which side of the gap this example fall in.

However, the performance of SVM heavily depends on the parameter selection. Based on the above analysis, we focus on the method of parameter selection for SVM in this paper. In this paper, we aim to detect salient objects in digital images with high accuracy and efficiency using modified SVM [5], and the local features are used to describe the visual content of images. Particularly, the innovations of this paper are mainly lie in the following aspects:

(1) We propose a novel approach to use a circular region to replace the rectangular region and we assume that the affine and scale normalized regions have been extracted in advance.

(2) To represent visual contents of the host image more accurately, we design a hybrid QPSO-SVM model through which the SIFT features clustering process and local features matching process can be implemented.

(3) To tackle the problem that the performance of SVM algorithm is greatly influenced by the quality of parameter selection, we utilize the quantum behaved particle swarm optimization (QPSO) technique to choose suitable SVM parameters through analyzing the convergence of particle swarm optimization and quantum system

(4) In the proposed method, salient objects can be detected efficiently by integrating QPSO and SVM together through image segmentation and interest points matching.

The rest of the paper is organized as the following sections. Section 2 introduces the related works of salient object detecting. Section 3 illustrates the proposed scheme for salient object recognition. In section 4, experiments are conducted to make performance evaluation with comparison to other existing methods. Finally, the whole paper is concluded in section 5.

II. RELATED WORKS

In this section, we will introduce the related works of this paper by three aspects. Firstly, salient objects extraction from a digital image is a very hot topic, the reasons lie in that salient objects include a lot of useful applications, such as image compression, content-based multimedia retrieval and so on.

Liu et al. proposed an approach to improve the performance of object extraction, and then illustrate a two phase salient objects extraction method by image segmentation technology and saliency detection process. Furthermore, in phase 1, the host images are segmented and then the saliency map for this image is distinguished through saliency detection algorithm. Moreover, for a given region, image visual features are extracted by Support Vector Machine to classify this chosen region as a background region of the host image [11].

Chen et al. proposes an efficient approach for automatic salient object detection, and then the authors made up a reliable saliency map to evaluate the image composition. Afterwards, the local energy feature is integrated with a simple biologically inspired model to promote the integrity of the object in the saliency map. Furthermore, pixels of minimal intensity are deleted from the host image and then the entropy of the resulting images are calculated [12].

Jing et al. propose a new threshold-free salient object detection algorithm combining saliency density and edge response. The authors illustrate a global optimal window which includes a salient object. This algorithm is efficiently located through the proposed saliency density and edge response based on branch-and-bound search. To detect the salient object with a well-defined boundary, the Grab Cut algorithm is implemented, and this algorithm is initialized through the located window [13].

Secondly, to effectively find salient objects in digital images local features have been widely utilized. As is well know that the detection and description of local features can help in the work of object detection. Particularly, SIFT features are local and based on the appearance of the object at particular interest points, and are invariant to image scale and rotation, and the following parts will illustrate the related works of applications of local features.

Zhang et al. illustrated a new image classification algorithm via Harr-like transformation of local features. In this paper, Harr-like transformation on local features is used to integrate the spatial information and the correlations of local features. Particularly, the Harr-like transformed local features are encoded utilizing the technology of non-negative sparse coding [14].

Bai et al. described an image fusion algorithm based on the extracted local image features through multi-scale top-hat by reconstruction operators. Firstly, the authors proposed a multi-scale local feature extraction method based on multi-scale top-hat and reconstruction operators. Secondly, the image visual features for image fusion are extracted by using the extracted multi-scale local features obtained from several images [15].

Li et al. proposed a new framework which is named iSearch, and tem global-local matching of local features are integrated together to promote the performance image retrieval quality. In this paper, Li et al. extracted multiple local features (such as scale-invariant feature transform, regional color moments and object contour fragments) to effectively represent the visual appearances of items. On the other hand, global and local matching of large-scale image dataset is permitted. Finally, the authors presented a contour fragments encoding and indexing algorithm [16].

Thirdly, Support Vector Machine algorithm has been used in object detection, because it can effectively detect the salient objects from the complex background from the large-scale digital images.

Zhao et al. proposed a new video object retrieval system framework. This system is design using a Spatio-Temporal data representation, a dedicated kernel design
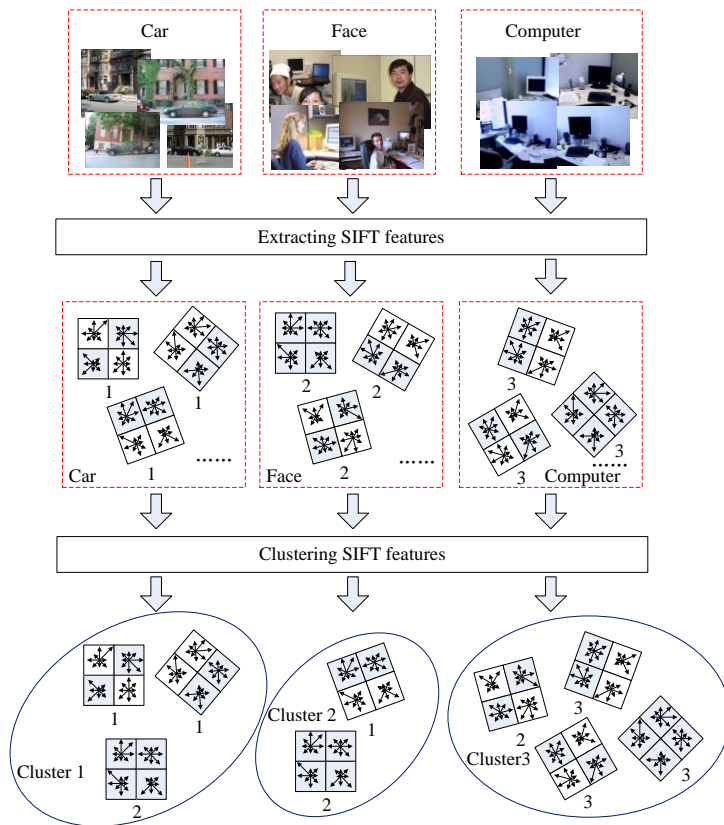
Figure 1.   Constructing the object classes from training image dataset using SIFT features

and a statistical learning toolbox. Using the kernel similarity, both supervised and active learning strategies embedded in Support Vector Machine framework are proposed in this paper. Furthermore, a multi-class classification framework dealing with unbalanced data is proposed as well [17].

Song et al. give six different voxel selection approaches to determine which voxels of fMRI data can be included in Support Vector Machine classifiers utilizing linear and RBF kernels. Next, the overall performances of voxel selection and classification methods are made comparison. Furthermore, this paper presented the first empirical result of linear and RBF Support Vector Machine in classification of fMRI data, integrated with voxel selection approach [18].

III.   PROPOSED SCHEME

As local features are more suitable to detect salient objects in images, we use SIFT descriptors to represent the interesting keypoints of the objects. The SIFT algorithm can extract features which are invariant to scaling, orientation, affine transforms and illumination changes. For each keypoint, users can calculate a local feature descriptor based on the local image gradient, transformed to compute orientation invariance. Afterwards, a SIFT feature vector can be modeled as the following equation.

$$V(F_i) = \left(\begin{bmatrix} v_1(F_i) \\ \cdots \\ v_{n_i}(F_i) \end{bmatrix}, \begin{bmatrix} loc_1(F_i) \\ \cdots \\ loc_{n_i}(F_i) \end{bmatrix} \right) \qquad (1)$$

where $v_1(F_i)$ denotes the feature vector of the keypoint F_i. Moreover, $loc_1(F_i)$ represents the location in the host image. It can be seen that the distance between vector $V(F_1)$ and $V(F_2)$ can not be calculated by Euclidean distance directly.

In this paper, we utilize a circular region to replace the rectangular region, because the rotational invariance performance of circular region is much better. Supposing that the affine and scale normalized regions have been extracted. To calculate the local descriptor, the derivatives I_{x1} and I_{x2} of the specific region $I(x)$ could be obtained as follows.

$$I_{x1}(x_1, x_2) = I(x_1, x_2 + 1) - I(x_1, x_2 - 1) \qquad (2)$$

$$I_{x2}(x_1, x_2) = I(x_1 + 1, x_2) - I(x_1 - 1, x_2) \qquad (3)$$

Afterwards, the magnitude and orientation of the host image can be calculated for a specific image region as follows.

$$M(x_1, x_2) = \sqrt{I_{x1}(x_1, x_2)^2, I_{x2}(x_1, x_2)^2} \qquad (4)$$

$$\theta(x_1, x_2) = \tan^{-1}\left(\frac{I_{x2}(x_1, x_2)}{I_{x1}(x_1, x_2)} \right) \qquad (5)$$

To implement the salient object detection, the object classes should be constructed from training image dataset using SIFT features in advance. As is shown in Fig. 1, several image of the object class "Car", "Face" and

"Computer" are chosen, and the SIFT features are extracted from the training images. Afterwards, these SIFT features are clustering. Furthermore, given a test image, after SIFT features extracting, the class which the test image belonged to can be found by SIFT features matching. In this paper, the SIFT features clustering process and local features matching are implemented through the following hybrid QPSO-SVM model.

To make the salient object detection more effectively, we use SVM to classify local features of the host images. Supposing there are a set of training images which are $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l)\}$, where the conditions $x_i \in R^d$ and $y_i \in R$ are satisfied. Given i, a regression model is defined which can describe the relation between x_i and $f(x_i)$:

$$f(x_i) = w \cdot x_i + b, b \in R, i \in \{1, 2, \cdots, l\} \quad (6)$$

To compute the parameter y_i for x_i, the constrained optimization problem should be calculated (shown in Eq.6).

$$\min_{w, b, \xi^{(*)}, \varepsilon} \tau(w, \xi^{(*)}, \varepsilon) = \frac{\|w\|^2}{2} + C \cdot \left(\frac{\sum_{i=1}^{l} (\xi_i + \xi_i^*)}{l} + v \cdot \varepsilon \right) \quad (7)$$

Subject to

$$(w \cdot x_i + b) - y_i \leq \varepsilon + \xi_i \quad (8)$$

$$y_i - (w \cdot x_i + b) - y_i \leq \varepsilon + \xi_i^* \quad (9)$$

where the conditions $\xi^{(*)} \geq 0$ and $\varepsilon \geq 0$ are satisfied. w and x_i represent d-dimensional column vectors. The value of v is belonged to the range $(0,1]$, and v is utilized to control the number of support vectors. Furthermore, ε is exploited to control the size of tube, and C refers to the trade-off value between separate margin and errors.

Moreover, parameter $\xi^{(*)}$ is equal to $(\xi_1, \xi_1^*, \xi_2, \xi_2^*, \cdots, \xi_l, \xi_l^*)^T$, which denotes the slack variables.

Particularly, for the linear separable problem, SVM algorithm is designed based on the following condition.

$$\begin{cases} w^T x_i + b \geq 1, & y = 1 \\ w^T x_i + b \leq -1, & y = -1 \end{cases} \quad (10)$$

Although SVM is a powerful computing tool, the quality of parameter selection is SVM could influence the system performance greatly. Therefore, in this section we will how to choose suitable SVM parameters using the quantum behaved particle swarm optimization(QPSO) technique, which is designed by analyzing the convergence of particle swarm optimization and quantum system. In QPSO, the state of quantum is represented as the function $\varphi(x, t)$, and exploiting the Monte Carlo

approach, the position of each particle is obtained by Eq. 11.

$$x_{ij}^{t+1} = p_{ij}^t + \chi \cdot L_{ij}^t \cdot \ln\left(\frac{1}{u_{ij}^t}\right) \quad (11)$$

$$L_{ij}^t = 2 \cdot \lambda \cdot \left| mbest_j^t - X_{ij}^t \right| \quad (12)$$

where the value of χ is equal to 0.5 or -0.5, and p_{ij}^t refers to the local attractor which is defined as follows.

$$P_{ij}^t = \varphi_{ij}^t \cdot P_{ij}^t + (1 - \varphi_{ij}^t) \cdot P_{gj}^t \quad (13)$$

where φ_{ij}^t represents a random number which is belonged in the range $(0,1)$, and P_{gj}^t refers to the global best position, and L_{ij}^t is calculated as follows.

$$L_{ij}^t = 2 \cdot \lambda \cdot \left| p_{ij}^t - X_{ij}^t \right| \quad (14)$$

In Eq. 14, parameter λ is used to adjust the convergence speed. Afterwards, the position can be updated by the following equation.

$$X_{ij}^{t+1} = p_{ij}^t + \lambda \cdot \left| p_{ij}^t - X_{ij}^t \right| \cdot \ln\left(\frac{1}{u_{ij}^t}\right) \quad (15)$$

The center of pbest positions for a specific swarm is computed by the following equation.

$$mbest^t = (mbest_1^t, mbest_2^t, \cdots, mbest_N^t)$$
$$= \left(\frac{\sum_{i=1}^{M} P_{i1}^t}{M}, \frac{\sum_{i=1}^{M} P_{i2}^t}{M}, \cdots, \frac{\sum_{i=1}^{M} P_{iN}^t}{M} \right) \quad (16)$$

Based on the above analysis, the position of particle could be updated as follows.

$$X_{ij}^{t+1} = p_{ij}^t + \lambda \cdot \left| mbest_j^t - X_{ij}^t \right| \cdot \ln\left(\frac{1}{u_{ij}^t}\right) \quad (17)$$

Combining the QPSO and SVM together, host images can be segmented effectively and the interest points can be matched accurately, and then salient objects are extracted from host images.

IV. EXPERIMENTS

In this paper, the a series of experiments are design and implemented to make performance evaluation using different datasets, and other related research works are compared with ours.

A. Datasets and Performance Evaluation Metric

The dataset used for salient object detection we chosen is provided from paper [19]. In paper [19], the task of salient object detection is implemented by finding a bounding box around the most salient object in the digital image. This dataset is made of five thousand images with

manually labelled rectangles near the most salient object drawn by different experts. To achieve the ground truth, we ask some experts to give a bounding rectangle to represent the object. The proposed salient object detection provides a rectangle on the salient object which is detected. Particularly, the ground truth is determined by select the rectangle around the salient object using the majority agreement of all the experts we chosen. The salient objects are represented by the bounding box with red color.

The dataset we used is made up of a large-scale image database with 130,099 digital images from several categories, and mostly of them are selected from the Web. Afterwards, we choose 60,000 digital images, in which each image includes at least one salient object. To make performance evaluation more objectively, another 20,840 images are selected which include an unambiguous object of interest to help for constructing the ground truth. Particularly, the salient objects we select are different in diversity, including "category", "color", "shape", "size" et al.. On the other hand, to show the performance of the proposed salient object detection approach, MSRC dataset [20] is also utilized which have many different object categories.

Furthermore, the performance evaluation metric we used are precision, recall, and F-measure. Precision/recall can represent the area ratio of detected object and the ground truth salient object region.

$$P = \frac{\sum_x g_x \cdot a_x}{\sum_x a_x} \qquad (18)$$

$$R = \frac{\sum_x g_x \cdot a_x}{\sum_x g_x} \qquad (19)$$

where P and R denotes Precision and Recall respectively. Afterwards, F-measure refers to the weighted harmonic mean of the metric Precision and Recall as follows.

$$F_\gamma = \frac{(1+\gamma) \times P \times R}{\gamma \times P + R} \qquad (20)$$

where F_γ denotes the F-measure with a noonegative parameter γ, and in this paper the parameter γ is equal to 0.5.

B. Experimental Resutls and Related Analysis

In order to make performance evaluation, other computing models are compared with the proposed SVM-QPSO model, such as SVM-PSO, SVM, ANN, and ARMA. SVM-PSO means integrating the standard SVM and Particle Swarm Optimization, and ANN and ARMA denotes artificial neural networks and Auto-Regressive and Moving Average Model respectively.

Particle Swarm Optimization(PSO) was designed by Eberhart and Kennedy, the idea of PSO lies in the social behavior of animals such as bird flocking, fish schooling, and the swarm theory and so on. Different from the genetic algorithm, PSO has some features and in many research fields and has been proved to be efficient. Artificial neural network(ANN) denotes the computational models inspired by animals' central nervous systems which are capable of machine learning and pattern recognition. Autoregressive moving average model(ARMA) represents a parsimonious description of a weakly stochastic process based on two polynomials, of which the first one is used for the auto-regression and the other one is used for the moving average.

As is shown in Table I, the proposed SVM-QPSO model performs better than other traditional methods when using the Precision, Recall and F-measure metric. Compared with ANN, ARMA, SVM and SVM-PSO, the F-measure of the proposed SVM-QPSO model promotes 3.4%, 4.24%, 21.02% and 9.42% respectively. Our SVM-QPSO model performs than other methods, because QPSO can choose the optimal parameters for SVM to promote the detection accuracy. Moreover, integrating SVM algorithm and QPSO has made the proposed SVM-QPSO model to perform better than other methods. The standard SVM performs poorly, and the reason lies in that the parameters of standard SVM could not be selected properly.

TABLE I.   PERFORMANCE EVALUATION FOR DIFFERENT METHODS UNDER THE METRIC PRECISION, RECALL, F-MEASURE

Category	ANN			ARMA			SVM			SVM-PSO			SVM-QPSO		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
tree	0.541	0.533	0.539	0.664	0.556	0.637	0.493	0.501	0.495	0.656	0.697	0.584	0.719	0.778	0.630
sign	0.579	0.674	0.608	0.687	0.587	0.654	0.519	0.531	0.523	0.750	0.629	0.602	0.677	0.727	0.637
flower	0.686	0.497	0.609	0.655	0.580	0.614	0.556	0.573	0.561	0.659	0.741	0.596	0.849	0.823	0.648
sheep	0.621	0.724	0.652	0.693	0.593	0.659	0.530	0.520	0.527	0.662	0.636	0.561	0.763	0.862	0.606
bird	0.659	0.620	0.646	0.584	0.610	0.534	0.492	0.543	0.508	0.751	0.611	0.592	0.837	0.811	0.669
aeroplane	0.562	0.641	0.586	0.625	0.619	0.563	0.518	0.486	0.507	0.716	0.630	0.600	0.757	0.859	0.680
body	0.550	0.679	0.587	0.676	0.548	0.628	0.497	0.515	0.503	0.721	0.578	0.591	0.889	0.725	0.656
car	0.697	0.516	0.624	0.621	0.594	0.591	0.493	0.605	0.526	0.646	0.846	0.578	0.722	0.906	0.605
building	0.618	0.743	0.655	0.672	0.539	0.649	0.553	0.585	0.563	0.766	0.800	0.597	0.726	0.759	0.662
cat	0.685	0.585	0.648	0.606	0.635	0.591	0.503	0.494	0.500	0.774	0.620	0.601	0.787	0.743	0.616
chair	0.650	0.667	0.656	0.669	0.530	0.598	0.547	0.565	0.553	0.795	0.702	0.588	0.673	0.714	0.617
cow	0.661	0.666	0.663	0.704	0.620	0.607	0.503	0.532	0.512	0.775	0.642	0.577	0.714	0.790	0.638
dog	0.643	0.524	0.598	0.699	0.615	0.618	0.572	0.587	0.577	0.707	0.731	0.567	0.856	0.805	0.705
face	0.507	0.627	0.542	0.670	0.587	0.611	0.522	0.528	0.524	0.708	0.594	0.575	0.754	0.892	0.616
book	0.669	0.649	0.662	0.704	0.618	0.645	0.530	0.556	0.539	0.695	0.626	0.558	0.750	0.859	0.600
Average	0.622	0.623	0.618	0.662	0.589	0.613	0.522	0.541	0.528	0.719	0.672	0.584	0.765	0.804	0.639

TABLE II.   EXAMPLES OF THE PROPOSED SALIENT OBJECT DETECTION ALGORITHM

p	Original image	Local feature extracting	Segmentation	Objects detected
No.1				
NO.2				
NO.3				
NO.4				
NO.5				
NO.6				

To illustrate the performance of each method more detailedly, the Precision-Recall curve are utilized, and the image categories "face", "car", "dog" and "book" are chosen(shown in Fig. 2-Fig. 5).
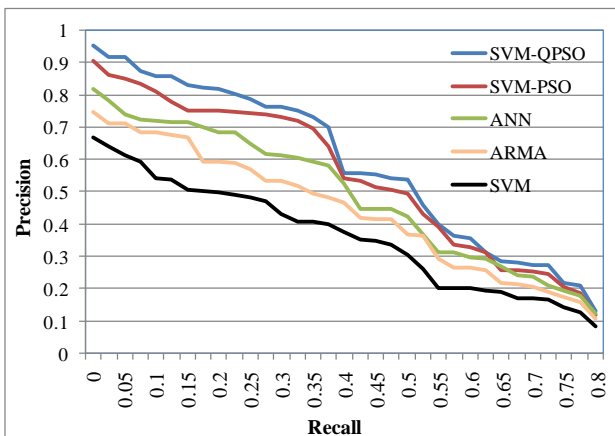
Figure 2.   Precision-Recall Curve for the category "face"
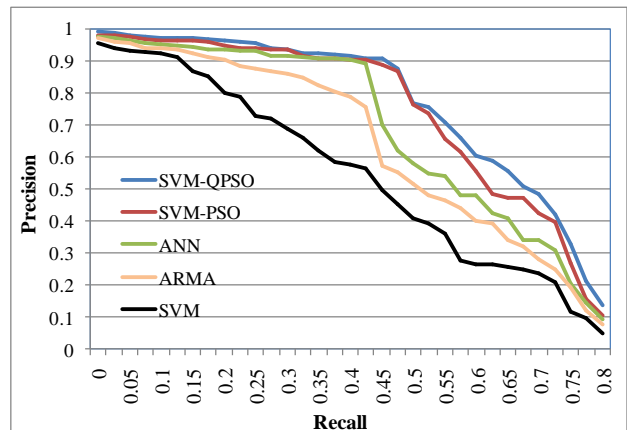
Figure 3.   Precision-Recall Curve for the category "car"

It can be seen from Fig. 2-Fig. 5 that the proposed salient object detection algorithm performs better that other four methods, especially for the category "car" and "dog", because the visual diversity of "car" and "dog" are less that other categories, and SIFT features is more

suitable for detect these categories. Although the salient object detecting precision is of great importance, the computing efficiency is a key issue as well. In the following section, we will test the computing efficiency for each algorithm.
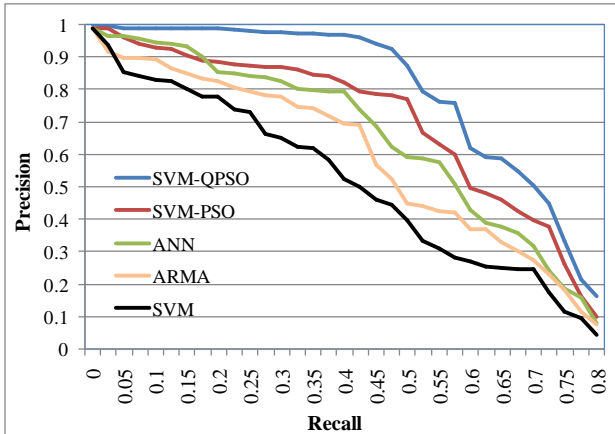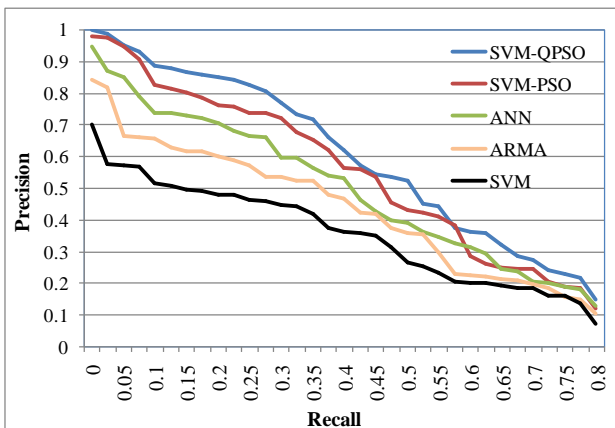

Figure 4.   Precision-Recall Curve for the category "dog"

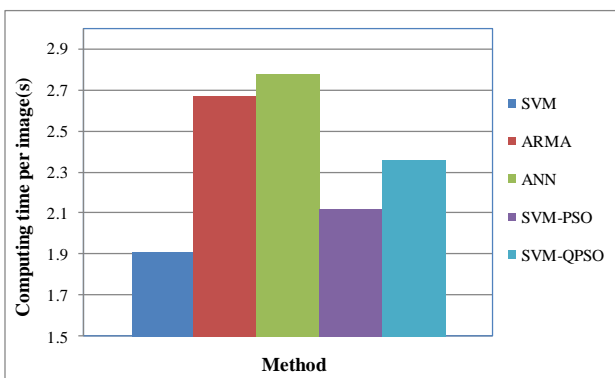

Figure 5.   Precision-Recall Curve for the category "book"

Figure 6.   Computing time for different methods

From Fig. 6, we can see that computing time per image of the proposed algorithm performs better than ARMA and ANN, however, less than SVM-PSO and standard SVM. However, salient object detection precision of SVM-PSO and standard SVM is much less than ours. Combining the all the above experimental results together, our proposed algorithm can effectively detect salient

objects both in detect precision and in computing efficiency.

From the examples in table II, we can see that the proposed algorithm can accurately locate the salient objects in host images through the red color bounding box, and the keypoints of each image are detected accurately by the SIFT features.

V.   CONCLUSIONS

In this paper, we propose a novel salient object recognition approach using local features. Firstly, the visual object classes are constructed from training image dataset via SIFT features clustering. Secondly, for a given test image, the object can be found through interest points matching. Thirdly, the SIFT features clustering and local features matching are achieved by the hybrid SVM-QPSO model. Fourthly, we exploit the quantum behaved particle swarm optimization technique to promote the accuracy of parameter selection for SVM.

REFERENCES

[1] Felzenszwalb Pedro F., Girshick Ross B., McAllester David, Object Detection with Discriminatively Trained Part-Based Models, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2010, 32(9) pp. 1627-1645
[2] Leibe Bastian, Leonardis Ales, Schiele Bernt, Robust object detection with interleaved categorization and segmentation, *International Journal Of Computer Vision*, 2008, 77(1-3) pp. 259-289
[3] Sheikh Y, Shah M, Bayesian modeling of dynamic scenes for object detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2005, 27(11) pp. 1778-1792
[4] Singh Navjot, Arya Rinki, Agrawal RK, A novel approach to combine features for salient object detection using constrained particle swarm optimization, *Pattern Recognition,* 2014, 47(4) pp. 1731-1739
[5] Du Juan, Chen Su-Li, Research of improved network malicious information filtering model based on SVM, *Journal of Networks,* 2013, 8(5) pp. 1012-1018
[6] Xue Jianru, Wang Le, Zheng, Nanning, Automatic salient object extraction with contextual cue and its applications to recognition and alpha matting, *Pattern Recognition,* 2013, 46(11) pp. 2874-2889
[7] Huang TieJun, Tian YongHong, Li Jia, Salient region detection and segmentation for general object recognition and image understanding, *Science China-information Sciences,* 2011, 54(12) pp. 2461-2470
[8] Kang DJ, Ha JE, Kweon IS, Fast object recognition using dynamic programming from combination of salient line groups, *Pattern Recognition,* 2003, 36(1) pp. 79-90
[9] Du Shuze, Chen Shifeng, Salient Object Detection via Random Forest, *IEEE Signal Processing Letters,* 2014, 21(1) pp. 51-54
[10] Li Hongliang, Meng Fanman, Ngan King Ngi, Co-Salient Object Detection From Multiple Images, *IEEE Transactions on Multimedia,* 2013, 15(8) pp. 1896-1909

[11] Liu Qiang, Han Tao, Sun Yantao, A two step salient objects extraction framework based on image segmentation and saliency detection, *Multimedia Tools And Applications,* 2013, 67(1) pp. 231-247

[12] Chen Xiao, Zhao Hongwei, Liu Pingping, Automatic salient object detection via maximum entropy estimation, *Optics Letters,* 2013, 38(10) pp. 1727-1729

[13] Jing Huiyun, Han Qi, He Xin, Saliency Density and Edge Response Based Salient Object Detection, *IEICE Transactions on Information and Systems,* 2013, E96D(5) pp. 1243-1246

[14] Zhang Chunjie, Liu Jing, Liang Chao, Image classification using Harr-like transformation of local features with coding residuals, *Signal Processing,* 2013, 93(8) pp. 2111-2118

[15] Bai Xiangzhi, Zhou Fugen, Xue Bindang, Image fusion through local feature extraction by using multi-scale top-hat by reconstruction operators, *OPTIK,* 2013, 124(18) pp. 3198-3203

[16] Li Haojie, Wang Xiaohui, Tang Jinhui, Combining global and local matching of multiple features for precise item image retrieval, *Multimedia Systems,* 2013, 19(1) pp. 37-49

[17] Zhao Shuji, Precioso Frederic, Cord Matthieu, Spatio-Temporal Tube data representation and Kernel design for SVM-based video object retrieval system, *Multimedia Tools and Applications,* 2012, 55(1) pp. 105-125

[18] Song Sutao, Zhan Zhichao, Long Zhiying, Comparative Study of SVM Methods Combined with Voxel Selection for Object Category Classification on fMRI Data, *PLOS ONE,* 2011, 6(2), No. e17191

[19] Liu, Tie and Yuan, Zejian and Sun, Jian and Wang, Jingdong and Zheng, Nanning and Tang, Xiaoou and Shum, Heung-Yeung, Learning to detect a salient object, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2011, 33(2) pp. 353-367

[20] J. Shotton, J. Winn, C. Rother, A. Criminisi, Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation, *The European Conference on Computer Vision,* 2006, pp. 1-15.

Simulation Study on Ventilation & Cooling for Main Transformer Room of an Indoor Substation

Tingfang Yu and Huijuan Yang

School of Mechanical and Electronic Engineering, Nanchang University, Nanchang, Jiangxi Province, China
Email: wtu_tingfy@163.com

Rui Xu

State Grid Jiangxi Electric Power Research Institute, Nanchang, China, 330006
Email: xurui@163.com

Chunhua Peng

School of Electrical & Electronics Engineering, East China Jiaotong University, Nanchang, Jiangxi Province, China
Email: pch_china@163.com

Abstract—**Based on natural ventilation design scheme for main transformer room of an indoor substation, different air distribution schemes were obtained by changing height and size of air inlets and outlets. Three-dimensional simulation of air distribution was conducted for the transformer room by using Computational Fluid Dynamics (CFD) method. Ventilation & cooling effect of different indoor ventilation schemes were simulated with software FLUENT. By analyzing velocity fields and temperature fields, influences of different design parameters on safety and reliability of main transformer room of indoor substation were compared and analyzed in details. Additionally, characteristics and change rules of air distribution with different parameter variations were concluded. Considerations of ventilation organization design for main transformer room of indoor substation and recommendation for better air distribution schemes were provided. The research results also offered some guidance for design and renovation of ventilation & cooling projects of transformer room.**

Index Terms—**CFD; Indoor Transformer Room; Ventilation & Cooling; Simulation Study**

I. INTRODUCTION

As electric power load of city center and difficulty of land acquisition became more and more outstanding problems in China, 110-220KV substations were more arranged in an indoor manner [1, 2]. Heat dissipation and ventilation of indoor transformer were key problems for rated output once the power transformers of high voltage and large capacity were equipped indoors. In summer, existing indoor substations in many cities had problems in ventilation and heat dissipation [3, 4]. Using empirical formula for design and model test method were two traditional ventilation design ways for indoor substations, however, the former one had poor computational accuracy and the later one would increase cost of labor and model for test equipment [2, 4]. In recent years, continuous improvement of numerical simulation method

and rapid development of computing technology brought more and more applications of Computational Fluid Dynamics (CFD for short) in the field of fluid computation design. Wu Ming [5] discussed impact of spacecraft flow field on stress and heat transfer of spacecraft was analyzed by utilizing CFD technology, and test time as well as design cost were reduced. Pan Sha [6] adopted CFD method to provide an effective way of seakeeping prediction for ship sailing in oblique sea. Zhou Jicheng [7] studied on the bubbly two-phase flow characteristics in twisted tube bundles, the method of CFD was used out to analyze the effects which angles between the major axis of the cylinder and vertical direction (θ) and bubble diameters have on the motion behaviours of bubbly flow. Cheng Wei-Min[8] performed research on eddy air-curtain dust controlled flow field in hard rock mechanized driving face, mathematical model of single-phase flow of gas was established based on the k-ε two-equation model and numerically simulated the eddy air-curtain dust controlled flow field in hard rock mechanized driving face with the help of FLUENT software. With advantages of low cost, non-contact fluid field measurement and availability of detailed flow field information, CFD was gradually applied in many other aspects [9].

In the research field of indoor ventilation and ventilation & cooling of indoor substation, Yi Jiang [10] and Tang Zhen-chao [11] respectively used CFD to analyze and calculate natural ventilation in buildings and conducted related study. With the aid of CFD method, relevant analysis and study were conducted on indoor ventilation & cooling of industrial building with residual heat in reference [12]. Regarding to CFD application in design and analysis of ventilation schemes of substation, related application studies were carried out to underground and semi-underground substations in reference [13-16]. While references [17, 18] conducted traditional analysis and comparison in aspects of ventilation & cooling of urban indoor substations, but

there were only a few studies using CFD numerical simulation means.

Xu Luwen [19] discussed a CFD (Computational Fluid Dynamics) numerical model of the main transformer, and analyzes the results with FLUENT software. The optimum heat dissipation effect with the optimum layout of the air inlet and outlet was obtained by changing the positions of the air inlet and outlet. T. vanHooff [20] Used the Computational Fluid Dynamics (CFD) to study complex physical processes in the built environment requires model validation by means of reduced-scale or full-scale experimental data. CFD studies of natural ventilation of buildings in urban areas should be validated concerning both the wind flow pattern around the buildings and the indoor airflow driven by wind and buoyancy. R. Ramponi [21] presents the follow-up study that focuses in more detail on validation with wind tunnel measurements and on the effects of physical and numerical diffusion on the cross-ventilation flow, validation of the coupled outdoor wind flow and indoor air flow simulations is performed based on Particle Image Velocimetry (PIV) measurements for four different building configurations.

In consideration of successful application of CFD in various fields, using CFD numerical simulation technology to carry out numerical simulation studies on ventilation & cooling air distribution for indoor substations and to assist optimization design of substations will be a new trend. With CFD method, model modifications and simulation tests can be conveniently carried out according to electrical equipment configurations, operation conditions, the climate, terrain and surroundings of substations, to obtain air flow fields and temperature fields of indoor substations under different design conditions and optimize ventilation & cooling organization schemes as well as verify the rationality of ventilation & cooling design schemes. Analysis bases for ventilation & cooling system design of new indoor substations and for ventilation & cooling optimization and renovation of substations in service could be provided.

In this study, Fluent, which is excellent CFD software, was adopted to analyze the velocity field and temperature field in the transformer room on the basis of natural ventilation design schemes. Ventilation characteristics of different locations and sizes of air inlets and air outlets were simulated. By changing boundary conditions of various design parameters, a total of 22 groups working conditions were simulated and calculated to obtain simulation results of each working condition. In this way, change rules of impact of design parameter variations on ventilation & cooling effect of the main transformer room in urban indoor substations were analyzed in accordance with simulation results.

The rest of the paper is organized as follows: Section II describes natural ventilation design for the transformer room, the parameters such as sizes and important temperatures are given in table I. Section III describes CFD numerical calculation model of the indoor substation, and boundary condition and parameter setting for numerical simulation calculation are given. Section IV discusses numerical simulation results and comparison analysis for different cases. Finally in Section V conclusions are given.

II. NATURAL VENTILATION DESIGN FOR THE TRANSFORMER ROOM

On the basis of Design Code for Substation (35kV-110kV) (GB50059-92) [23], ventilation design was calculated for main transformer room of 110kV substation. Design structural parameters of the transformer room: the dimension of main transformer room was 10m×10m×12m (length×width×height), the dimension of transformer was 5.42m×4.92m×5.63m (length×width×height), and there were 16 sets of radiators arranged in the same room with the transformer located in the center with a low arrangement. The lowest location was 0.2m from the ground and was 2.29m from the left and right walls and 2.54m from the front and back walls. Figure 1 showed detailed structure of transformer room of the indoor substation.
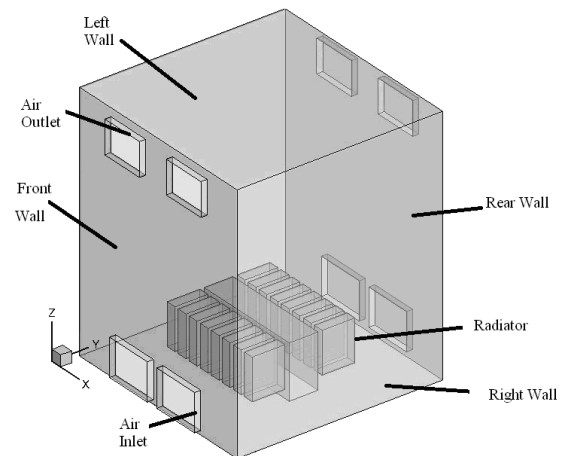

Figure 1.   Structure of transformer room

The natural ventilation design was based on thermal pressure ventilation and ventilation design of the transformer room was closely related to operation parameters of the transformer. As stipulated in the national standard of China [23], for ventilation in Summer, calculated temperature of the transformer room should not exceed 35 ℃, air exhaust temperature should below 45 ℃ and temperature difference between inlet and outlet should not exceed 15 ℃. The original design parameters were as below: for ventilation in Summer, the design temperature was 30 ℃, the volume of ventilation and heat dissipation was 179.61kW and the air exhaust temperature was 44 ℃.

Parameters such as sizes and locations of air inlets and outlets, ventilation volume, temperature in working area and air exhaust temperature could be obtained in natural ventilation design, see Table I for details.

Check results basing on measured data and air inlet temperature of 30 ℃ showed that the natural ventilation design could satisfy requirements of design standard; see Table II for detailed check results.

TABLE I.   RESULTS SUMMARY OF NATURAL VENTILATION DESIGN

Item	Calculated temperature in Summer / ℃	Height difference between the air inlet and outlet /m	Location of neutral plane /m	Dimension of the air inlet /m	Number of air inlets	Dimension of the air outlet
Value	33	10	3.8	2.4×1.8	4	2.2×1.6
Item	Number of air outlets	Area ratio of the air inlet and outlet	Air exhaust temperature / ℃	m value	Average indoor temperature / ℃	Ventilation volume kg/s
Value	4	1.24	44	0.6	42	17.78

TABLE II.   RESULTS SUMMARY OF CHECKING CALCULATION

Check item	Inlet air temperature / ℃	Location of neutral plane /m	Air exhaust temperature / ℃	Ventilation volume kg/s
Checking result	30	3.8	39	19.8

TABLE III.   THE SIMULATION RESULTS OF DIFFERENT OUTLET ELEVATION

Elevation	C(8.8m)	B(9.8m)	A(design)	D(roof)
Air exhaust temperature /K	314.4	313.73	312.64	312.65
Air exhaust volume /kg/s	7.81	8.34	9.28	9.18
Average temperature of upper working area /K	311.96	311.21	309.99	309.68
Average temperature of working area /K	304.44	304.42	304.41	304.42

Notes: air exhaust parameters were average parameters of cross section of air outlet and the working area refers to the area within 2.6m above the ground
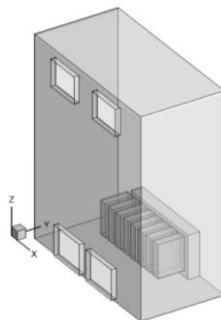

Figure 2.   The calculation model

III.   ESTABLISHMENT OF NUMERICAL CALCULATION MODEL

A. Simulation Object Analysis

In consideration of the symmetrical structure of the main transformer room and calculation amount, only half of the physical computational model established was considered here. Simplification was made to the transformer and radiator as the transformer was complicated, see Figure 2 for three-dimensional diagram of the computational model. As shown in the figure, there were 4 air inlets symmetrically arranged in front and back walls, the size of the air inlet was 2.4m×1.8m×0.3m, area of air inlet was 17.7m^2, the distance between center lines was 3m and the lower edge was 0.2m from the bottom. There were 4 air outlets symmetrically arranged with 2 arranged in the front wall and 2 in the back wall, the size of the air outlet was 2.2m×1.6m×0.3m, the distance between center lines was 3.8m and the center line was 1.2m from the top.

Figure 2 was the 3D physical model established with Gambit. Hexahedral grid was used to divide grids with grid spacing of 0.1-0.2m.

B. Mathematical Model

Air ventilation and heat transfer problem of the main transformer in indoor substation is the object of our study, the flow of air complied with the following basic equation:

a) Continuity equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} + \frac{\partial \rho w}{\partial z} = 0 \qquad (1)$$

b) Navier-Stokes equations:

$$\frac{\partial (\rho u)}{\partial t} + \frac{\partial (\rho u^2)}{\partial x} + \frac{\partial (\rho uv)}{\partial y} + \frac{\partial (\rho uw)}{\partial z}$$
$$= -\frac{\partial p}{\partial x} + \mu \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) + \rho F_x \qquad (a)$$

$$\frac{\partial (\rho v)}{\partial t} + \frac{\partial (\rho uv)}{\partial x} + \frac{\partial (\rho v^2)}{\partial y} + \frac{\partial (\rho vw)}{\partial z} \qquad (2)$$
$$= -\frac{\partial p}{\partial x} + \mu \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \right) + \rho F_y \qquad (b)$$

$$\frac{\partial (\rho w)}{\partial t} + \frac{\partial (\rho uw)}{\partial x} + \frac{\partial (\rho vw)}{\partial y} + \frac{\partial (\rho w^2)}{\partial z}$$
$$= -\frac{\partial p}{\partial x} + \mu \left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} \right) + \rho F_z \qquad (c)$$

where in, ρ is the fluid density, u, v, w is the velocity component of the fluid at the point of (x, y, z) under time of t, p is pressure, μ is the fluid dynamic viscosity, F_x, F_y, F_z are the mass force of x, y, z direction respectively .

c) Energy conservation equation

$$\frac{\partial (\rho T)}{\partial t} + \frac{\partial (\rho uT)}{\partial x} + \frac{\partial (\rho vT)}{\partial y} + \frac{\partial (\rho wT)}{\partial z}$$
$$= \frac{\partial}{\partial x} \left(\frac{\lambda}{c_p} \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left(\frac{\lambda}{c_p} \frac{\partial T}{\partial y} \right) + \frac{\partial}{\partial z} \left(\frac{\lambda}{c_p} \frac{\partial T}{\partial z} \right) + S_h \qquad (3)$$

where in T is temperature, λ is the thermal conductivity of the fluid, C_p is the specific heat capacity of fluid, S_k for the entry of the fluid within the heat source.
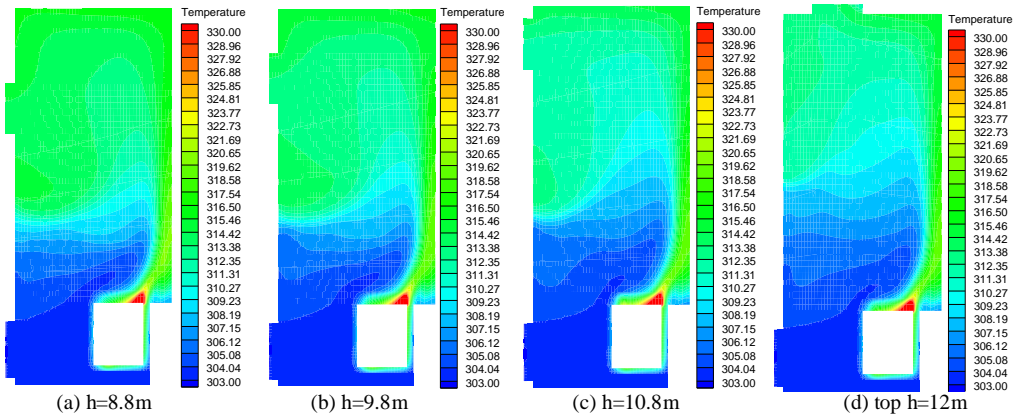
(a) h=8.8m   (b) h=9.8m   (c) h=10.8m   (d) top h=12m

Figure 3. Comparison of temperature field of x =3.4m cross section under different air out elevations

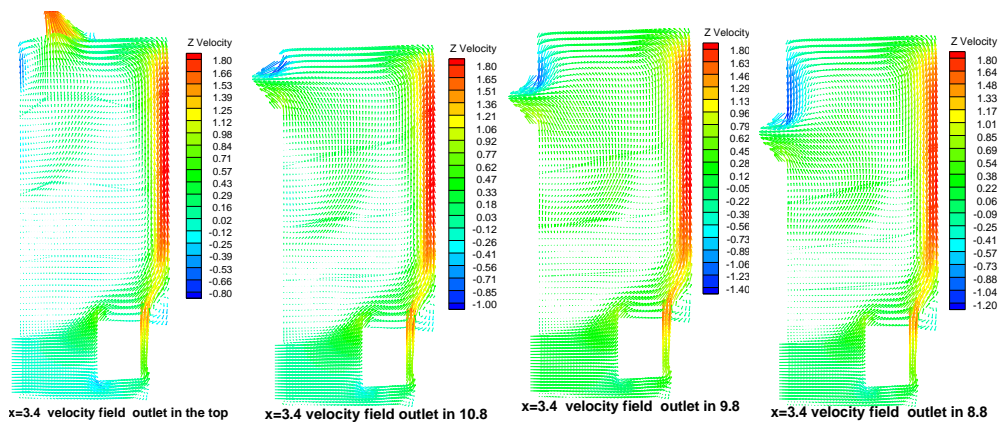x=3.4 velocity field outlet in the top   x=3.4 velocity field outlet in 10.8   x=3.4 velocity field outlet in 9.8   x=3.4 velocity field outlet in 8.8

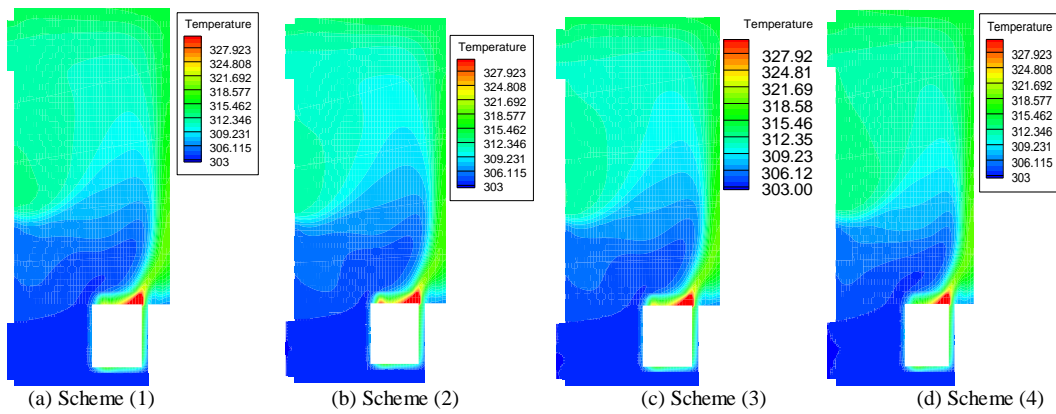Figure 4. Comparison of velocity vector distribution of x =3.4m cross section under different air out elevations

(a) Scheme (1)   (b) Scheme (2)   (c) Scheme (3)   (d) Scheme (4)

Figure 5. Comparison of temperature field of x=3.4m cross section under different air outlet area

(a) h=0.4m   (b) h=1.2m   c) h=2.6m
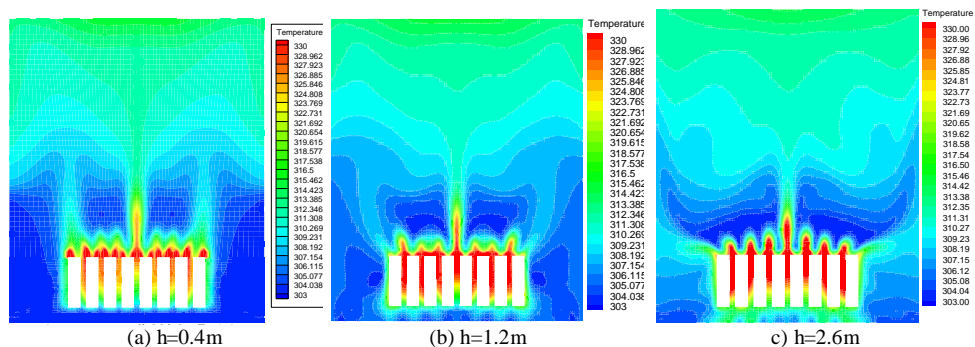
Figure 6. Comparison of Temperature Field of Y=4m under Different Inlet Elevations

TABLE IV.   NUMERICAL SIMULATION RESULTS OF DIFFERENT AIR OUTLET AREAS

Air outlet cross section	2.2m×1.4m	2.2m×1.6m	2.4m×1.6m	2.2m×1.8m
Air exhaust volume /kg/s	8.795	9.280	9.678	9.989
Air exhaust temperature /K	313.146	312.641	312.203	311.867
Average temperature of working area / ℃	304.413	304.415	304.395	304.386

Notes: air exhaust parameters were average parameters of cross section of air outlet and the working area refers to the area within 2.6m above the ground

TABLE V.   THE SIMULATION RESULTS IN DIFFERENT INLET AREA

Width (m)× height (m)	Air exhaust volume /kg/s	Air exhaust temperature /K	Average temperature of upper working area /K	Average temperature of the working area /K
1.8m×1.8m	9.246	312.637	310.802	304.922
2.0m×1.8m	9.252	312.630	310.473	304.693
2.4m×1.6m	9.251	312.645	310.246	304.540
2.4m×1.8m	9.280	312.641	309.990	304.415
2.6m×1.8m	9.283	312.618	309.828	304.349
2.4m×2.0m	9.321	312.577	309.817	304.335
2.8m×1.8m	9.282	312.617	309.735	304.260
2.4m×2.2m	9.339	312.560	309.672	304.284

The CFD simulations are performed with the 3D steady Reynolds-Averaged Navier–Stokes (RANS) approach with the standard k-ε turbulence model to provide closure [22]. As follows:

d)Turbulent kinetic energy (TKE, k):

$$\frac{\partial k \overline{u_i}}{\partial x_i} = \frac{\partial}{\partial x_i}(\frac{v_t}{\sigma_k}\frac{\partial x_k}{\partial x_i}) + v(\frac{\partial \overline{u_i}}{\partial x_j} + \frac{\partial \overline{u_j}}{\partial x_i})\frac{\partial \overline{u_j}}{\partial x_i} - \varepsilon \quad (4)$$

e) Energy dissipation rate (ε):

$$\frac{\partial \varepsilon \overline{u_i}}{\partial x_i} = \frac{\partial}{\partial x_i}(\frac{v_t}{\sigma_k}\frac{\partial k}{\partial x_i}) + C_{1\varepsilon}\frac{\varepsilon}{k}v(\frac{\partial \overline{u_i}}{\partial x_j} + \frac{\partial \overline{u_j}}{\partial x_i})\frac{\partial \overline{u_j}}{\partial x_i} - C_{2\varepsilon}\frac{\varepsilon^2}{k} \quad (5)$$

The turbulent viscosity (v_t) is calculated in terms of k and ε by:

$$v_t = C_\mu \frac{k^2}{\varepsilon} \quad (6)$$

where in the modeling constants $C_{1\varepsilon}$,$C_{2\varepsilon}$, C_μ ,σ_k , σ_s are 1.44, 1.92, 0.09, 1.0 and 1.3, respectively.

C. Boundary Condition and Parameter Setting

Indoor air flow was considered as 3D stable-state flow in this article and the driving force was thermal buoyancy (i.e. natural convection) [24, 25]. In calculation, 3D stable-state solution basing on separation solution was applied, and standard k-ε turbulence model and Boussinesq model were used to treat buoyancy. SIMPLEC velocity coupling method was adopted. In addition, the default relaxation factor was used due to flow of strong body forces. For pressure term, body force weighted schemes was used and the second upwind scheme was used for remain spatial discretization schemes to increase accuracy of the result.

Boundary condition: to be conforming to the principle of thermal pressure ventilation, pressure inlet of the air inlet and pressure outlet of the air outlet were adopted. For other boundaries, wall boundary condition was used.

Heat flux of the radiator wall was 1122w/m^2 and other walls were adiabatic boundaries.

IV.   NUMERICAL SIMULATION RESULTS AND ANALYSIS

To analyze the influence of sizes and locations of air inlets and outlets on ventilation & cooling effect of the transformer room, 22 schemes were developed in this article, including ventilation & cooling schemes of changing height and size of air inlets and outlets. Characteristics of ventilation air distribution and change rules of ventilation & cooling effect under various parameter variations were analyzed in a detailed manner.

A. Impact of Air Outlet Elevation on Ventilation Effect

To compare impacts of air outlet elevation on indoor ventilation & cooling, four types of air outlet elevation were simulated, including design parameters of 10.8m (A), elevation 9.8m (B), elevation 8.8m (C) and building roof elevation 12m (D). The numerical simulation results are listed in Table III and the temperature field and velocity field are shown in Figures 3 and 4.

As data in Table III shows, in the case of arranging air outlets at sides of walls (front and back walls), the air exhaust volume would increase and the air exhaust temperature would reduce as elevation increased. The main reason was thermal pressure, which was the main driving force of natural ventilation, increased as elevation increased, thus air exhaust volume increased. It showed that height increase would improve indoor ventilation & cooling effect.

The temperature field in Figure 3 and the velocity field in Figure 4 showed that: 1) elevation reduction of air outlets largely impacted the velocity field of the top area. Local recirculation zone formed at the top due to high temperature air flow, which caused obvious differences in temperature fields of the top area. The lower the air outlet elevation was, the higher the overall temperature of the top area would be; 2) it was mainly controlled by air inlet parameters in the working area, thermal plumes were not intensively mixed, temperature of the working area was therefore low; 3) neutral plane, which was a thermal

separation layer, existed at a certain height. The layer increased as the air outlet location rose. Distinct temperature gradient existed below the thermal separation layer, while the temperature gradient above the layer was small.

Resistance of high temperature air flow elimination reduced as air outlet height increased, which facilitated air flow elimination and was conducive to indoor ventilation & cooling. Under the current building structure, indoor high temperature air would form vortexes when the air outlet was arranged too low, which would increase overall temperature in indoor working area and enlarge the high temperature area. As a result, air outlets should be arranged as high as possible during design.

B. Impact of Air Outlet Area on Ventilation Effect

To obtain the impact of air outlet area on ventilation effect, working conditions under 4 different air outlet areas (including the design scheme) were simulated. Corresponding air outlet sizes are shown as follows:

(1) Design parameter 2.2m×1.6m
(2) Cross section size 2.2m×1.8m
(3) Cross section size 2.4m×1.6m
(4) Cross section size 2.2m×1.4m

Numerical simulation results of different areas are shown in Table IV and Figure 5. As Table 4 shows, air exhaust volume increased obviously as the area of air outlet increased, as a result, temperature of indoor parts reduced. Additionally, air flow elimination resistance reduced due to the increased area, residence time of high temperature air flow in the room therefore reduced. Vortexes around the air ports were weakened, which largely facilitated the improvement of indoor temperature distribution.

It is observed from temperature field of X=3.4m cross section in Figure 5 that, temperature fields under different air outlet areas differed obviously. Indoor high temperature areas reduced as the air outlet area increased. Downward backflow of high temperature air flow at the top was weakened and distinct difference existed between temperature fields in the top area. Impact of air outlet area on temperature fields in working area was small. As the air outlet area increased, rising of thermal plumes would increase.

However, in actual projects, air outlets were in the type of ventilation skylights. Larger air outlets cost more and noise suppression effect on the main transformer room would reduce, thus outdoor noise would exceed limits. This was an important consideration in practice.

C. Impact of Air Inlet Elevation on Ventilation Effect

To analyze the impact of air inlet elevation on indoor ventilation & cooling, 9 different air inlet elevations (height of air inlets above the ground) were simulated in this article. Heights (m) of elevations above the ground were within the range of 0.2m-2.6m. For simulation results of temperature fields and numerical simulation trends, please refer to Figure 6 and 7 respectively.

Figure 6 is the comparison of temperature field of Y=4m cross section under different inlet elevations. High

temperature areas around radiator in the working area enlarged as air inlet elevation increased. Vortexes of large area emerged and high temperature area in the working area expanded obviously. The overall temperature field distribution was in a mess and sections existed in temperature distribution, which was bad for heat discharge.

As shown in Figure 7 (temperature variation curve), temperature in the working area increased as air inlet elevation increased, and temperature in upper of the working area and the air exhaust temperature reduced slightly. This suggested that air inlet elevation had a significant influence on lower temperature field and no great impact on upper temperature field. Furthermore, the figure shows that, temperature field variation was small when air inlet elevation varied below 1m. Generally speaking, preferable ventilation & cooling effect in the working area could be achieved if air inlets were arranged below 1m, and temperature distribution would be worse and unfavorable to indoor ventilation if air inlets were above 1m.

D. Impact of Air Inlet Area on Ventilation Effect

To verify the impact of air inlet area on ventilation effect, working conditions under 8 different air inlet areas (including design scheme) were simulated. See Table V for parameters of relevant air inlet cross section.

As data of Table V and temperature variation curve in Figure 8 show, air exhaust volume would increase upon expansion of air inlet area, and indoor temperature would reduce accordingly, however, as the area enlarged, ventilation effect improvement would be small. It showed that corresponding effect could only be achieved when air inlet area matched with the air outlets. As seen from lines in the figure, increase of cooling effect would be smaller when the area reached 2.6m×1.8m.
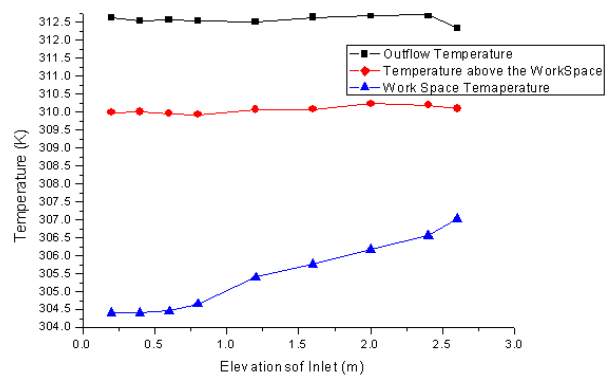

Figure 7.   Variation curve of indoor temperature parameters under different inlet elevations

Figure 9 is the comparison of temperature field of Y=4m cross section under different air inlet elevations. Velocity of inlet air flow would reduce as the air inlet area increased, which would rise the temperature in the radiator area. As turbulence intensity of inlet flow reduced and disturbance weakened, entrainment of thermal plumes increased and high temperature air flow diffusion in the central area would be weakened, which

would be effective for temperature field improvement in other areas. It showed that increase of air inlet area had some effect on temperature field improvement.
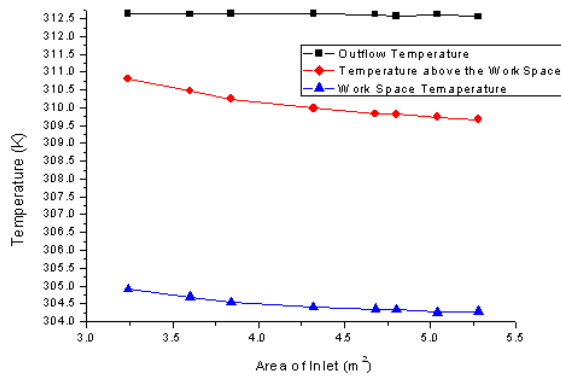

Figure 8.   Temperature variation curve under different air inlet areas
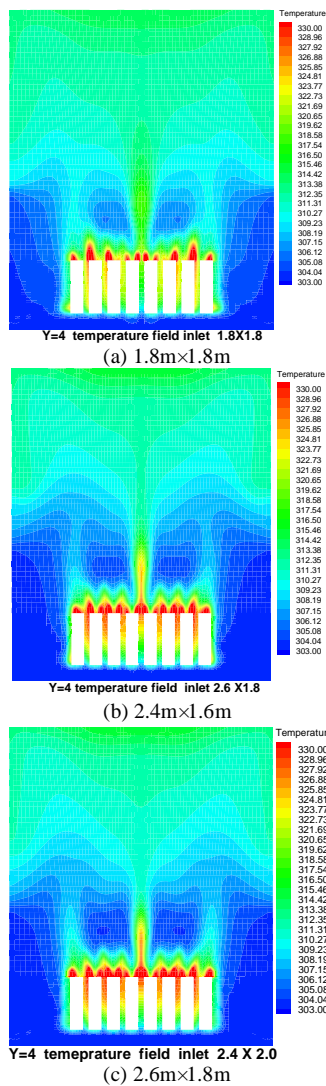

(a) 1.8m×1.8m

(b) 2.4m×1.6m

(c) 2.6m×1.8m

Figure 9.   Comparison of temperature field of Y=4m cross section under different air inlet areas

Comparison of Figure 9 (b) and (c) shows that reducing the high temperature area in the lower part by increasing area in the width could better improve the temperature field in the lower area and facilitate heat dissipation of the transformer.

V.   SUMMARIES

CFD technology was applied in ventilation schemes design for air distribution of main transformer room of indoor substation. Basing on the design scheme, a total of 22 groups of working conditions were simulated and calculated by changing boundary conditions of various design parameters, and simulation results of each working condition was obtained. According to the results, following conclusions were drawn by analyzing ventilation & cooling effect change rules of main transformer room of urban indoor substation under different design parameters.

(1) This ventilation & cooling design scheme for main transformer room of the indoor substation was reasonable and could meet the requirements of design standards.

(2) Numerical simulation results of the ventilation design applying CFD calculation was consistent with the design calculation, and the calculation accuracy met the design standard. Tests before design and cost of real model were greatly reduced. Simulation results could provide references for ventilation & cooling system design of new indoor substations and for ventilation & cooling optimization and renovation of substations in service.

(3) According to comprehensive rules analysis basing on simulation results, heat pressure could be increased to improve indoor temperature and velocity fields and to enhance ventilation & cooling effect by increasing height differences of the air inlets and outlets on the premise of building structural condition and noise control; it is appropriate to arrange air inlets below 1m and ventilation volume could be enhanced by increasing area of air inlets and outlets, in which way temperature and velocity fields could be improved.

REFERENCES

[1] Zhang Guizhu, Developing trends of Substation in big cities, *China Electric Power Education*, vol. 12, pp. 90-91, 2011

[2] Chai Ting, On air-conditioner ventilation design scheme for some110kV substation, *ShanXi Architecture,* 2011, (3): pp. 126-127

[3] JIN Linfang. Research on Ventilation and Noise Reduction of Main Transformer Chamber in Urban Indoor Substation. *Electricity Supply and Consumption*, 2010, (4): pp. 39-41

[4] Han Wenqing, Gan L. 220 kV Ventilation system design for indoor 220 kV substations. *Heat Ventilation & Air Condition*, 2011, (5): pp 47-49

[5] Wu Ming, Shi Aiguo, Yang Bo. Simulation research on three DOF motions of ship in oblique waves based on CFD. *Application Research of Computers*, vol. 30, pp. 2233-2235, 2013

[6] Pan Sha, Li Hua, Xia Zhi-xun. High-Performance Parallel Computation Application for Aerospace CFD Numerical Simulation. *Computer Engineering and Science,* vol. 34, pp. 191-198, 2012

[7] Zhou Jicheng, Zhu Dongsheng. Study on bubbly two-phase flow across twisted tube bundles based on quasi 3D high speed video. *Journal of Multimedia*, vol. 8, n 6, pp. 823-829, 2013

[8] Cheng Wei-Min, Research on eddy air-curtain dust controlled flow field in hard rock mechanized driving face. *Journal of Networks,* vol. 8, n 2, pp. 453-460, February 2013

[9] Peng, Zhou. Implementation of optimal pacing scheme in xinjiang's oil and gas pipeline leak monitoring network, *Journal of Networks*, vol. 6, n 1, pp. 54-61, 2011

[10] Yi Jiang, Qinyan Chen. Study of natural ventilation in building by large eddy simulation. *Journal of wind engineering and industrial aerodynamics*, vol. 89, pp. 1155-1178., 2001

[11] Tang Zhen-chao, Zhan Jie-ming. Application of CFD to the Research of Air Flow in Room. *17th National Symposium hydrodynamic hydrodynamic Sixth National Conference Proceedings*, 2003:8.

[12] Wan Xin. Numerieal Simulation of Natural Ventilation in Industrial Plant with Heat source. *Mater Dissertation, Donghua University*, 2008.

[13] Juan Carlos Ramos, Maximiliano Beiza, Jon Gastelurrutia, Alejandro Rivas, Raúl Antón, Gorka S. Larraona, Iván de Miguel, Numerical modeling of the natural ventilation of underground transformer substations, *Applied Thermal Engineering,* Vol. 51, Issues 1–2, pp. 852-863, March 2013.

[14] YANG Xiao-ping, LIU Gang, YANG Qi-cai. Numerical Simulation on Cooling and Ventilation of Transformer Room in Underground Transformer Substation. *Building Energy&Environment*, 2011, (1) pp. 94-96

[15] HUANG Qiang, LIU Xin, CHEN Lei. Study on Air Distribution in Transformer Plants of Substations of Metro Stations: Case Study on a Station of Tianjin Metro. *Tunnel Construction*, 2010, (4): pp. 396-401

[16] HU Wen-bin, LIU Li-na. Discussion of ventilation and air flow distribution in semi-underground transformer substation. *Journal of Harbin University of Commerce (Natural Sciences Edition)*, 2011, (3): pp. 346-349

[17] GUO Sheng, HUANG Wei, XUE Lin. Ventilation mode of indoor substation based on integrated environmental control technology. *Journal Of Beijing Jiaotong University*, 2013, (3): pp. 19-26

[18] SHU Li-fan, HU Wen-bin, CHEN Ming-lan, et al. Research on Hybrid Ventilation Mode of Indoor Substation. *Building Science*, 2012, (2): pp. 84-87

[19] XU Luwen, WANG Guochao, DING Jun et al. The CFD Numerical Simulation on Ventilation of Indoor Transformer Substation. *Hydromechatronics Engineering*, Vol. 41, No. 18: pp. 141-144., 2013

[20] T. vanHooff, B. Blocken. Full-scale measurements of indoor environmental conditions and natural ventilation in a large semi-enclosed stadium: Possibilities and limitations for CFD validation. *Journal of Wind Engineering and Industrial Aerodynamics*, 2012, (104-106): pp. 330-341.

[21] R. Ramponi, B. Blocken. CFD simulation of cross-ventilation flow for different isolated building configurations: Validation with wind tunnel measurements and analysis of physical and numerical diffusion effects. *Journal of Wind Engineering and Industrial Aerodynamics,* 104–106 (2012) : pp. 408–418

[22] James O. P. Cheung, Chun-Ho Liu. CFD simulations of natural ventilation behaviour in high-rise buildings in regular and staggered arrangements at various spacings. *Energy and Buildings,* vol. 43, pp. 1149－1158, 2011

[23] PRC Ministry of Construction. GB50059－92 35~110kv Substation design specifications.

[24] Chen Yu Xu Zhihao Ma Guochuan. The Study about CFD Algorithms of Natural Ventilation. *Refrigeration and Air Conditioning*, 2011, (1): pp. 78-81

[25] Duan Shuangping, Zhang Guoqiang, Peng Jianguo, etal. Development in research of natural ventilation, *Heat Ventilation & Air Condition*, vol. 34, pp. 22-28, 2004

Tingfang Yu received the B. Tech degree in thermal power engineering from Shanghai University of Electric Power, Shanghai, China, 1991 and the M.S. and Ph.D. degrees in thermal power engineering from the Southeast University, Nanjin , China, in 1998 and 2004, respectively.

He is currently an Associate Professor of Thermal Power Engineering with the School of Mechanical and electronic Engineering, Nanchang University, Nanchang, China. He has authored or coauthored over 20 research papers in journals and conferences, His current research interests are in the field of coal-fired boiler combustion and optimization algorithm

Huijuan Yang received the B. Tech degree in Thermal Power Engineering from, Nanchang University, Nanchang, China, 2013. He is currently a graduate student of Thermal Power Engineering in Nanchang University, Nanchang, China.

Handwriting Digital Recognition via Modified Logistic Regression

Cungang Wang

School of Computer Science, Liaocheng University, Shandong, China
Email: c.g.wanglc@gmail.com

Abstract—**Motivated by a wide range of real world applications of hand writing digital recognition, e.g., postal code recognition, the past decades have seen its great progress. The related approaches are generally composed of two components, feature extraction and identification methods. We note that the previous approaches are limited by the following two aspects: (1) the feature is not adaptive enough to cover the great variance within data; (2) the recognition methods are suffered from local minima solution. Inspired by these observations and to overcome these limitations, we in this paper propose an approach HMM-MLR by exploiting hidden Markov model (HMM) and modified logistic regression (MLR). In the proposed approach, HMM is employed to model the trace of handwriting digital, which is able to model the large variance within digitals and can adapt to the data distribution. Then the features are extracted based on HMM and then delivered into MLR for recognition. Benefitting from the global optimum solution of MLR, the proposed approach could reach highly stable results. To verify the effectiveness of the proposed approach, we experimentally compare our proposed approach with other state-of-the-art approaches over Semeion handwritten digit dataset. The experimental results show that, over both recognition accuracy and recall, for different rounds of experiments and different number of training samples, our HMM-MLR exhibits significant improvement over others.**

Index Terms—**Handwriting Digital Recognition; Hidden Markov Model; Modified Logistic Regression**

I. INTRODUCTION

Researchers started the investigation of pattern recognition theory and applications since 1920s. Along with the rapid development of computer and artificial intelligence techniques, it has been a branch of science until early 1960s. The continuous development of recognition techniques [1] obtains a wide range of applications, for instance, character recognition, image processing, speech recognition, biosensor. Among them, handwriting digital recognition and image recognition [2] are important applications, on the base of image processing, image representation, pattern recognition techniques. The progress of these techniques brings more and more researchers to this field, working on both algorithms and applications [1, 2].On the other hand, according to the statistics, for human being, more than 70% information is obtained through visual signal, which indicates that visual signal is the most important way for

communication. The large amount of images or visual information has been more and more important during the object recognition in recent years [1]. For this reason, image recognition becomes a popular topic [1-3].

In recent years, the fast development of artificial intelligence field provides a number of new approaches [1, 2] especially for image recognition and handwriting digital recognition [2]. As an important tool for understanding the real world, recognition techniques play an increasingly important role in both industry and daily life. With the development and applications of Internet and information science, the scale of image is increasing. Specifically, handwriting digital recognition techniques are used to handle more 10 millions of mails every data. Moreover, handwriting digital recognition techniques are demanded by character recognition is document analysis.

Researchers introduced a number of methods [1, 6-10] into image recognition [1, 5], such as neural network, support vector machine, Bayesian decision theory, hidden Markov models (HMM) [11]. These methods can be categorized into two classes, discriminative learning approaches and generative learning approaches. Typical discriminative approaches include support vector machine, neutral network, naïve Bayes, which typical generative approaches include HMM and so on.

Artificial neural network is based on empirical risk minimization and suffers from over-fitting, slow learning speed and other disadvantages during the pattern recognition. For support vector machine (SVM) [5], its model parameters exert a tremendous influence to the identification performance [6,7]. That is, appropriate parameters tend to have high performance of image recognition and handwriting digital recognition while inappropriate parameters cause degeneration in both computational efficiency and recognition accuracy. The naive Bayesian method could hardly satisfy the large scale identification problem. And it is difficult to realize the assumption of recognition unconditional independence. Additionally, it increases the learning and training complexity.

A stochastically process is named as Markov process if its future state at time $t+1$ only depends on its current state at time t, and is independent with its past states before time t. Hidden Markov Model (HMM) is a model on the base of the Markov assumption, with hidden states introduced. The observed events are not conducted deterministically. And they were connected

using a set of probability distribution. HMM was a dual random process, which consisted of two randomly processes, Markov chain as well as general randomly process. Markov chain described the state transition by means of transition probability, while general stochastically process described the relationship between state and observation series through observation probability. HMM [11] is a probability model which describes the statistic property of stochastically process by parameters. The model has been widely adopted in the areas of speech identification, character recognition, genetic analysis as well as bioinformatics engineering, etc.

From the above disadvantages, this paper proposes an approach for handwriting digital recognition that is on the basis of the hidden Markov model (HMM) and modified logistic regression (MLR). First, we enhance and thin the image of handwriting digit, so that we are able to extract the skeleton of digitals and then are able to separate them into individual digital. Second, we extract the location points from the skeleton and quantize the sequence of location points to a sequence of direction vectors. The reason is that, the direction vector is robust to rotation of digital which location sequence does not. Third, we use HMM to model the distribution of direction sequence of digitals. This can be implemented using standard learning method. Forth, the transition matrix conditioned on sample is employed as feature to represent the sample. Such feature exploits the data distribution and hidden states. Fifth, the extracted feature is delivered into MLR for recognition.

The contributions of the proposed method in this paper are threefold: (1) we propose to model the handwriting digit using hidden Markov model (HMM) and extract features based on HMM. To our best knowledge, the proposed approach is novel to this field. (2) We propose a modification for logistic regression so that it has better generalization ability than logistic regression. The modified logistic regression (MLR) is computationally efficient. (3) The proposed approach HMM-MLR combines the abilities of HMM and MLR for a comprehensive recognition [12].

The remainder of this work is organized as follows. Section 2 will present the proposed method, including hidden Markov model for feature extraction, modified logistic regression for recognition. Section 3 will verify the proposed method by a group of carefully designed experiments. We draw a conclusion in Section 4.

II.   OUR PROPOSED SCHEMA

On the basis of the above discussion, in this section, we propose a method for handwritten digit recognition. Our approach is based on the hidden Markov model and modified logistic regression, and is able to surmount the limitation of previous approach. Our developed algorithm is graphically illustrated in Fig. 1. There are three main steps. First, collect data; second, feature extraction via HMM model in this Section; third, MLR model training and test [13].
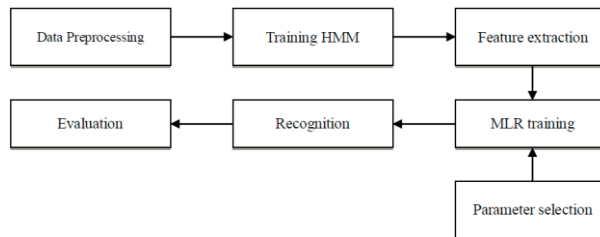

Figure 1.   The flow chart for experiment

A. Modeling Handwritten Digit Using HMM

The hidden Markov model (HMM) is a probabilistic generative model developed to model the distribution of sequences. It is composed of a set of hidden states and a set of observed states which are controlled by the hidden states. Here we first introduce the mathematical notations and then present the formulation of hidden Markov model. The quintuple $\lambda = (N, O, A, B, \pi)$ is employed to describe HMM [11], where $O = \{V_1, V_2, \ldots, V_M\}$ is the set of M observed states; $S = \{S_1, S_2, \cdots, S_N\}$ is the set of N hidden states; o_t is the observed state at time t; q_t is the hidden state at time t.

Given the above notations, we now introduce the basic probabilities. Let $A = \{\alpha_{ij}\}$ be state transition probability matrix, where α_{ij} is transition probability from state S_i to state S_j. Let $B = \{b_{kj}\}$ be the output probability matrix where b_{jk} is the output probability from hidden state S_j to observed state V_k. Let $\pi = \{\pi_i\}$ be the initial probability where $\pi_i = P(q_1 = S_i)$. When solving an estimation problem, for a given model and state transition series $q = (q_1, q_2, q_3 \ldots q_T)$, the probability of an observed series $O = (o_1, o_2, o_3 \ldots o_T)$ could be computed through the following equation:

$$P(o|q, \lambda) = P(o_1|q_1) \cdots P(o_T|q_T) = b_{q_1}(o_1) \cdots b_{q_T}(o_T) \qquad (1)$$

If λ is given, the probability of $q = (q_1, q_2, q_3 \ldots q_T)$ can be compute through

$$P(q|\lambda) = \pi_{q_1} \alpha_{q_1 q_2} \cdots \alpha_{q_{T-1} q_T}$$

Therefore, the estimation probability of the observed sequence can be is written as,

$$P(o, q|\lambda) = P(o|q, \lambda) P(q|\lambda) \qquad (2)$$

An observation sequence may have more than one corresponding state transition sequences; thereby all the state transition sequences could be is expressed as

$$P(o|\lambda) = \sum_q P(o, q|\lambda)$$
$$= \sum_{q1, q2, \ldots qT} \pi_q b_{q1}(o_1) a_{q1 q2} b_{q2}(o_2) \cdots a_{qT-1 qT} b_{qT}(o_T) \qquad (3)$$

The hidden Markov model can be efficiently solved using forward-backward algorithm [11].

B. Modified Logistic Regression for Identification

Both real and binary responses could adopt Logistic Regression, whose output posterior probabilities can be processed expediently and sent to other systems. It attempts to simulate the class label's conditional probability offer its observation:

$$p(y \mid x) = \frac{1}{1 + \exp\left(-y\left(w^T x + b\right)\right)} \qquad (4)$$

where $x = (x_1, \cdots, x_m)^T$ is the example vector; m is the number of features; $y \in \{+1, -1\}$ is the category label; $w = (w_1, \cdots, w_m)^T$ is the weight vector; b is decision intercept. The weight can be estimated as,

$$\hat{w} = \arg\min_w \left\{ \frac{1}{n} \sum_{i=1}^n \log\left[1 + \exp\left(-y_i\left(w^T x_i + b\right)\right)\right] + \lambda w^T w \right\}$$

Please notice which the Hessian matrix of the objective function $O(w)$ is:

$$H = \frac{d^2 O(w)}{dw \, dw^T} = \frac{1}{n} \sum_{i=1}^n \frac{\exp\left(-y_i w^T x_i\right)}{\left(1 + \exp\left(-y_i w^T x_i\right)\right)^2} x_i x_i^T + 2\lambda I$$

Here the identity matrix is I. Thanks to $\lambda > 0$, the Hessian matrix above mentioned is positive definite, that means the objective function of regularized LR with strict convexity, hence, we can see its solution, which is unique and global.

In this part, we first state that, by constructing a series of optimization problems, the solutions converge to the solution of SVM. Thus, SVM could be solved by simple unconstrained optimization techniques. Then we put forward our simple MLR-CG algorithm that uses CG as its inner loop.

To simplify our derivations, we take advantage of the augmented weight vector $w = (b, w_1, w_2, \ldots, w_m)^T$ and the augmented sample vector $x = (1, x_1, x_2, \ldots, x_m)^T$ from right now unless otherwise specified. To keep the SVM optimization problem unchanged, its form becomes,

$$\hat{w} = \arg\min_w \left\{ \frac{1}{n} \sum_{i=1}^n \max\left\{0, 1 - y_i w^T x_i\right\} + \lambda \sum_{j=1}^m w_j^2 \right\}$$

The intercept $\omega_0 = b$ is not in the regularization term. We also need not to penalize the intercept ω_0 in the regularized LR to approximate SVM:

$$\hat{w} = \arg\min_w \left\{ \frac{1}{n} \sum_{i=1}^n \log\left(1 + \exp\left(-y_i w^T x_i\right)\right) + \lambda \sum_{j=1}^m w_j^2 \right\}$$

From former discussions we could see which loss functions play an important role in the SVM and LR. The SVM loss function can be approximated by the loss of the following modified LR:

$$g_\gamma(x, y, w) = \frac{1}{\gamma} \ln\left(1 + \exp\left(-\gamma\left(y w^T x - 1\right)\right)\right) \quad (5)$$

If we could approximate the SVM loss function,

$$g_{svm}(x, y, w) = \max\left\{0, 1 - y w^T x\right\}$$

With the above sequence of functions $\{g_\gamma\}$, then the problem can be resolved with simple unconstrained optimization techniques.

The conjugate gradient method follows the above convergence proof. That is to say, we calculate the solution of SVM by solving the problems a sequence of sub-optimization problems. Especially, we solve each sub-optimization problem by means of the conjugate gradient. For solving large-scale nonlinear optimization problems, conjugate gradient is one of the most popular methods. More importantly, it is compared with other methods in fitting LR, which shows that it is more efficient. We could determine that the HS direction in the experiment is more efficient than the other two directions. We make a list of our conjugate gradient method below, which is an iterative method in the inner loop [14].

Algorithm 1: Conjugate Gradient
1. Initialize $w = 0$, $\gamma = 1.0$, $l = 10$, and $\delta = 10$
2. Repeat until convergence:
(a) Initialize conjugate gradient by collocation its search direction to minus gradient
(b) Minimize O_γ with l steps
(c) Increase $\gamma \leftarrow \gamma + \delta$

In fact, we should start from small γ as well as run not increase γ to infinity for two reasons. The first reason is which when γ is big the Hessian matrix is ill-conditioned, that will result in the unsuitability of our algorithm. Starting from small γ and gradually increase it will lead us obtain the stable solution.

For example, when $\gamma = 200$, it is at most 0.003, that is already not influential for our problems. As well as later on, in our experiments we will give a summarize that this approximation will not degrade the performance of our trained classifier. We run not need increase γ after each conjugate gradient procedure since we should let conjugate gradient do at least several steps to fully adopt its power in finding conjugate directions; and also we do not require to wait until it converged before we change γ. We set both δ as well as l to be 10 in our experiments. And each round when γ is changed, conjugate gradient should be re-initialized. We use 200 conjugate gradient steps in our experiments as the stopping criteria. And we also make use of other criteria like the change of weight vector or objective value.

III. EXPERIMENTAL RESULTS

In this section, we will validate our proposed method HMM-MLR for handwritten digit recognition. The experimental steps contain the following steps: (1) data collection; (2) model the data and extract the feature; (3) train the model and perform test. The experimental procedure are summarize in the above section. This part will sequentially report the dataset, verification criterion and experimental results [15].

A. Experimental Database

The Simenon Handwritten Digit instance used in our

experiments were collected using Tactile Sal, Brescia, Italy and donated in 1994 to Simenon Research Center of Sciences of Communication, Rome, Italy, for machine learning research. 1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular box 16×16 in a gray scale of 256 values. Then each pixel of each image was scaled into a binary value by a fixed threshold. Each person wrote on a paper all the digits from 0 to 9, twice. The commitment was to write the digit the first trial in the normal way (trying to write each digit accurately) as well as the second time in a fast way (with no accuracy). The best validation protocol for this database seems to be a 5×2CV, 50% Tune (Train +Test) and completely blind 50% Validation. This dataset consists of 1593 records and 256 attributes. Each record represents a handwritten digit, originally scanned with a resolution of 256 grays scale (28). Each pixel of the each original scanned image was first stretched, as well as after scaled between 0 and 1 (collocation to 0 every pixel whose value was under the value 127 of the grey scale as well as setting to 1 each pixel whose value in the grey scale was over 127). Each binary image was scaled again into a 16×16 square box (the final 256 binary attributes). The categories described as shown in Table I.

TABLE I.   CLASS DISTRIBUTION OF SAMPLES

Categories	Number of samples
Man	873
Woman	720
Total number of samples	1593

B. Verification Criterion

To validate the ability of our proposed HMM-MLR method for handwritten digit recognition, as well as compare with different approach, we in this paper use the following assessment criterion to assess the HMM-MLR approach for handwritten digit recognition. Specifically, we define the recognition accuracy, recognition precision, identification recall and classification true negative rate. The definitions are report in Table II. In this table, TN represents true negative; TP denotes true positive; FN represents false negative; FP denotes false positive. These evaluation standards can be directly made use of to measure two class recognition problem of handwritten digit recognition, and multiple class recognition problem of handwritten digit recognition

TABLE II.   THE EVALUATION CRITERION THE HMM-MLR APPROACH FOR HANDWRITTEN DIGIT RECOGNITION

Evaluation criterion	Definition
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$
Precision	$TP/(TP+FP)$
Recall	$TP/(FN+TP)$
True negative rate	$TN/(TN+TP)$

C. Main Results

In the first experiment, we validate our proposed HMM-MLR approach for handwritten digit recognition, over the Semeion Handwritten Digit dataset. We employ two comprehensive criterions, accuracy and recall, for experimental verification. Identification accuracy and

recall are two typical and popular measures for the correctness of the identification model. The experimental procedure is show in the experiment section. The preprocessing step as well as feature extraction step is important due to their encoding discriminant information. Our proposed algorithm HMM-MLR is trained by the above approach, and some parameters of HMM-MLR are obtained using cross-validation strategy. We conduct the test for multiple rounds, where in each round we stochastically divide the database to training configure as well as test configure.

We extensively compare our proposed HMM-MLR method for handwritten digit recognition with three algorithm, HMM-KNN, HMM as well as SVM. The classification results are show in Table III as well as Fig. 2. These experimental results indicate which: (1) our proposed method HMM-MLR outperforms all three compared methods significantly, under the distinct experimental configurations, distinct number of training example, and distinct evaluation standard. (2) Our developed method exhibit robustness against the round of experiments, and the verification criterion, which no wonder mean which our proposed algorithm could be used to a lot of tasks. The reasons are three folds. (1) The HMM-MLR has the ability to map the nonlinear data in the low dimensional space to the high dimensional space by the feature extraction, which makes the classification problem simple. (2) In comparison with empirical parameter selection approach, the selection method for parameters can adapt to the dataset. (3) The experimental procedure of our method could provide informative features and could maximize the discrimination ability.

TABLE III.   CLASSIFICATION PERFORMANCE COMPARISION OF FOUR METHOD FOR HANDWRITTEN DIGIT RECOGNITION

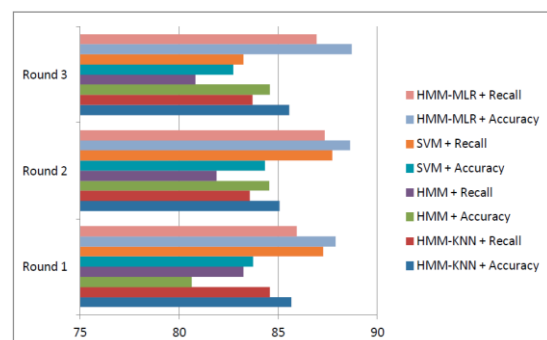| Experiment round	approach	Evaluation Criterion	
		Accuracy	Recall
Round 1	HMM-KNN	85.65	84.57
	HMM	80.62	83.23
	SVM	83.72	87.27
	HMM-MLR(ours)	87.88	85.93
Round 2	HMM-KNN	85.05	83.55
	HMM	84.55	81.88
	SVM	84.32	87.71
	HMM-MLR(ours)	88.62	87.35
Round 3	HMM-KNN	85.56	83.69
	HMM	84.57	80.81
	SVM	82.73	83.25
	HMM-MLR(ours)	88.70	86.93

Figure 2.   The performance comparison of experimented algorithm for handwritten digit recognition

In the second experiment, we validate the ability of the proposed HMM-MLR algorithm in handwritten digit recognition, by means of comparison experiment. Two popular criterions accuracy as well as recallis adopt for evaluation. Identification accuracy and recall are two typical and popular measures for the correctness of the identification model. The used database is Semeion Handwritten Digit dataset. The experimental step is show in above part. Our algorithm HMM-MLR is learnt by means of the method in above part, where some parameters of HMM-MLR are configure to defaults. The tests are run for several rounds over stochastically separate dataset.

The accuracy and recall is use as the assessment standard for the handwritten digit recognition. We do the experiment for 20 rounds as well as summarize the experimental results of partial round are in Table IV as well as Fig. 3. As report in Table IV, using by our method to determine the parameters, HMM-MLR for handwritten digit recognition reaches the highest performance of 91.29% under the standard of accuracy, while HMM-MLR achieve the highest performance of 88.45% under the standard of recall. Moreover, the average accuracy of HMM-MLR is 88.70% that outperforms which of HMM-KNN (85.43%). The potential reasons for these results are mainly threefold. Firstly, HMM-MLR has the ability to map the nonlinear data points in the low dimensional space to the high dimensional space by HMM based feature extraction, which simplifies the classification problem. Secondly, the parameter selection can be adaptive to different dataset, in comparison with the empirical parameter selection method. Thirdly, the framework of the proposed approach is composed of a group of comprehensive procedures which sequentially maximize the identification ability.

TABLE IV.   THE PERFORMANCE COMPARISON OF DIFFERENT METHOD

Experiment	Algorithm	Evaluation Criterion	
		Accuracy	Recall
Round1	HMM-KNN	85.65	84.57
	HMM-MLR(ours)	87.88	85.93
Round 2	HMM-KNN	85.05	83.55
	HMM-MLR(ours)	88.62	87.35
Round 3	HMM-KNN	85.56	83.69
	HMM-MLR(ours)	86.78	87.96
Round 4	HMM-KNN	84.47	83.29
	HMM-MLR(ours)	88.19	86.10
Round 5	HMM-KNN	84.61	83.75
	HMM-MLR(ours)	91.29	86.83
Average	HMM-KNN	85.43	84.83
	HMM-MLR(ours)	88.70	86.93

In the third experiment, we conduct experiments over Semeion Handwritten Digit dataset. The dataset includes have 1583 samples with 256 attributes, and the dataset is divided into two part, the training data set with 1400 samples and the test data set with 183 samples. This experiment verifies the ability of the proposed HMM-MLR for handwritten digit recognition, and the capability of the optimization method. The experimental scheme is summarize in the previous part of this work. It adopts a criterion approach to learn the HMM-MLR

based on the database Semeion Handwritten Digit dataset. In this experiment, HMM-MLR adopts the default parameters configure by means of authors. The assessment criterions are accuracy as well as recall where Identification accuracy and recall are two typical and popular measures for the correctness of the identification model.
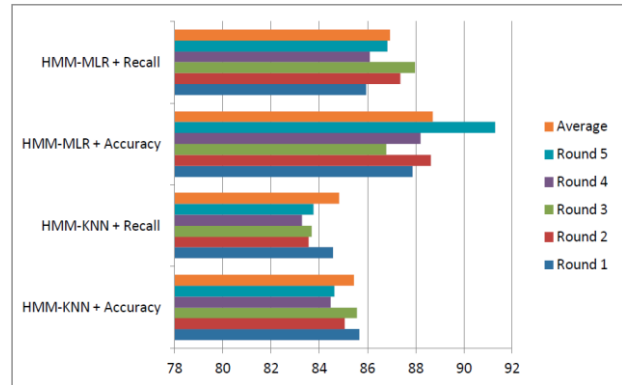

Figure 3.   The comparison of experimented method over two criterion

TABLE V.   THE PERFORMANCE COMPARISON OF HMM-MLR FOR HANDWRITTEN DIGIT RECOGNITION

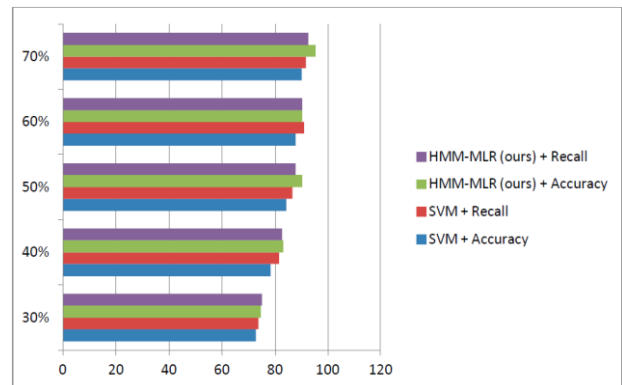| Training Samples	Method	Evaluation Criterion	
		Accuracy	Recall
30%	SVM	72.85	73.83
	HMM-MLR (ours)	74.57	75.07
40%	SVM	78.44	81.49
	HMM-MLR (ours)	83.08	82.77
50%	SVM	84.44	86.66
	HMM-MLR (ours)	90.26	87.87
60%	SVM	87.87	91.04
	HMM-MLR (ours)	90.26	90.36
70%	SVM	90.12	91.71
	HMM-MLR (ours)	95.41	92.59

Figure 4.   The performance comparison of HMM-MLR for handwritten digit recognition

Here we verify our developed algorithm by comparing it with HMM over two criterions, accuracy as well as recall. The experimental results for handwritten digit recognition are summarized in Table V as well as Fig. 4. As summarize in this table, we can see that, the values of accuracy as well as recall of our developed algorithm are much higher than that in HMM. That is to say, about 90.26-84.44% for accuracy and 87.87-86.66% for recall. We also found that, HMM-MLR is higher than HMM for

distinct number of training data, over two criterions. In all experiments with different collocation, the results of our approach exhibit advantage over the HMM approach. These results can be intuitively explained. First, comparing with the traditional machine learning based methods the HMM-MLR can be well applied to the condition that the sample data is large scale, high dimension and contains a large number of heterogeneous information. Second, in comparison with the empirical parameter selection method, our method can adapt to the dataset. Third, the framework of the proposed algorithm contains a group of comprehensive procedures which sequentially maximize the performance.

IV. SUMMARY

This paper proposed a handwriting digital recognition method based on hidden Markov model and modified logistic regression. Specifically, hidden Markov model models the distribution of sequence of digital, where the digital is preprocessed and represented as a sequence of direction vector. Then, the features of handwriting digital are extracted from hidden Markov model, i.e. use the matrix of transfer state as the feature. The extracted features are then delivered to modified logistic regression for recognition. The modified logistic regression method is an extension of logistic regression by modifying its loss function which approaches to the hinge-loss of support vector machine, but keeps the global optimum solution. The proposed approach HMM-MLR has the follows three advantages: (1) the sequence of direction vector is robust to the rotation of images; (2) the extract feature is adaptive to data distribution and exploits hidden states; (3) the modified logistic regression has global optimum solution which produces robust results. The experiment results over real world database, by comparing with other popular methods validate the advantages of our proposed approach.

REFERENCES

[1] Duda, Richard O and Hart, Peter E and Stork, David G. Pattern classification. *John Wiley & Sons,* 2012.
[2] Haijun Li, Application Research of Food Quality Detection Based on Computer Vision, *Journal of Computers,* Vol 8, No. 7, pp. 1758-1762, 2013
[3] Xu, Lei and Krzyzak, Adam and Suen, Ching Y. Methods of combining multiple classifiers and their applications to handwriting recognition, *IEEE Transactions on Systems, Man and Cybernetics,* vol 22, no 3, pp. 418-435, 1992.
[4] Huang Weiyong, Shao Xiaogen, Chen Kui, Coal-and-Gas Outburst Forecast Using CCPSO and SVM, *Computer Science,* Vol. 39, No. 11, pp. 216-220, 225, 2012
[5] Haralick R M. Statistical and Structural Approaches to Texture, *Proceeding of IEEE,* Vol. 67, No. 5, 1957
[6] Tang Jingtian, Hu Dan. SAR image classification research based on SVM. *Remote Sensing Technology and Application,* Vol. 23, No. 3, pp. 56-58, 2008
[7] Li Yi, Huachun Zhou, Fei Ren, Hongke Zhang, "Analysis of Route Optimization Mechanism for Distributed Mobility Management", *Journal of Networks,* Vol. 7, No. 10, pp. 1662-1669, 2012
[8] Yi Yan, Zhao Tao, Qi lili. Evaluation and research of knowledge management maturity based onneural network of particle swarm. *Journal of Information,* Vol. 29, No. 5, pp. 124-128, 2010
[9] Jian Wu, Zhi-ming Cui, Jian-ming Chen, Guang-ming Zhang, A Survey on Video-based Vehicle Behavior Analysis Algorithms *Journal of Multimedia,* Vol. 7, No. 3, pp. 223-230, 2012
[10] Jing Li, Xiaowen Zhang Geometrically Invariant Watermarking Scheme Based on Local Feature Points, *Journal of Multimedia,* Vol. 7, No. 3, pp. 231-238, 2012
[11] Guangming Zhang, Zhiming Cui, Pengpeng Zhao, JianWu, A Novel De-noising Model Based on Independent Component Analysis and Beamlet Transform, *Journal of Multimedia,* Vol. 7, No. 3, 247-253, 2012
[12] Eddy, Sean R, Hidden markov models, *Current opinion in structural biology,* vol. 6, no. 3, pp. 361-365, 1996.
[13] Yueqiu Jiang, Xiangwen Fu, Hongwei Gao, A New Circular Region Detection Algorithm based on the Geometric Characteristics, *Journal of Software,* Vol. 8, No. 11, pp. 2899-2907, 2013
[14] Wei Zhan, Zhi-Qing Luo, Research of Vehicle Type Recognition System Based on Audio Video Interleaved Flow for Toll Station, *Journal of Software,* Vol. 7, No. 4, pp. 741-744, 2012
[15] Zhenbing Liu, Shujie Jiang, Huihua Yang, Jianguo Liu, Arbitrary-length Fast Hartley Transform without Multiplications, *Journal of Computers,* Vol. 8, No. 10, pp. 2656-2663, 2013

Cungang Wang received his B.E. degree in educational technology from Liaocheng University, Liaocheng, China, in 2000, and his M.S. degree in computer science and technology from Ocean University of China, Qingdao, China, in 2006. He worked as a lecturer at the School of Computer Science at Liaocheng University since 2008. His research interests include machine learning, image processing and pattern recognition.

A Local Similarity Pattern for Removal of Random Valued Impulse Noise

Shan Jianhua and Zhu Liangliang
Department of Mechanical Engineering, Anhui University of Technology, Ma'anshan City, China
Email: 379751793@qq.com, 469151691@qq.com

Abstract—**This letter presents a new filtering scheme based on local similarity pattern within the local window for removing random valued impulse noise. A pixel to be considered as an original pixel, it should have ample numbers of similar neighboring pixels in a local window. The neighbors are divided into 2 subtypes: smooth similar neighbors and edge similar neighbors according to the different criterion. Extensive simulations show that the proposed filter provides better performance than many of the existing filters. In particular, the thresholds are adaptive to diverse image types at different noise rate and the computational complexity is very low.**

Index Terms—**Image Filtering; Random-Valued Impulse Noise; Impulse Detection**

I. INTRODUCTION

Digital images are often corrupted by impulse noise during the acquisition or transmission processes. The corrupted images severely impede subsequent image processing, such as image segmentation, edge detection or object recognition. Therefore, a fundamental problem in image processing is to suppress impulse noise from an image, whereas keeping the original pixels intact and preserving its details. The impulse noise can be classified as fixed-valued and random-valued impulse noise. In this paper, we focus on random-valued impulse noise which is uniformly distributed in the dynamic range of [0, 255] and which is more difficult to identify.

The median filter, which exploits the rank-order information of pixel intensities within a filtering window and replaces the center pixel with the median, is the most popular choice for removing impulse noise from images because of its effectiveness and simplicity. Nevertheless, it tends to distort fine details, since it modifies both noisy as well as noise-free pixels. Consequently, many filtering algorithms have been proposed which use some thresholds to detect noisy pixels, such as Progressive median (PSM) [1], MSM [2], tri-state median (TSM) filter [3], EM [4] and the pixel-wise MAD (PWMAD) filter [5] and, etc [6-8]. The performance of these filters is heavily dependent on the capabilities of the impulse detectors. Unfortunately, these filters exhibit poor performance for heavily corrupted noisy images. The reason is that they detect noisy pixels across the whole image region without considering the difference between the locally smooth region and edge region. The intensity difference pattern between the center pixel and other pixels in the filtering

window is not the same for the locally smoothing region and edge region, so they need different criterion to detect noisy pixels. In order to overcome this drawback, in [9, 10] image rank ordered absolute differences (ROAD) is proposed to measure the similarity or closeness of a pixel value to its neighbor, which utilizes the sort information and threshold value to determine the noisy-free pixels. Due to the high complexity of sort operation, ROAD is limited in application.

In [11] using different directional filter convolution to detect the noisy-free pixels, the optimal thresholds are hard to select.

The [12, 13] use fuzzy mathematics for noise removal, the problem of this method is hard to select the right fuzzy functions and de-fuzzy function. Recently, many new methods, such as NASMBF [14], UTMF [15], ABDND [16] ASMF [17] and NSC [18], [19-21] are some other schemes which are proposed for detection noise and some of its variants. Even if great progresses have been made, there is still room to improve.

In this letter, a new scheme based on local similarity patten in the local window is presented, which exhibits significantly improved impulse detection capability whereas very simple and fast. Impulse detection has respective criterion for smooth and edge region. Most of impulse detection available in the literature are performed in an iterative manner and need time-consume sort operation and varying filter window size. Our method does not need these.

The letter is structured as follows. In Section II, we introduce the draft detection scheme and the recovery process. In Section III we show numerical results and visual examples. Finally, Section IV provides the concluding remarks.

II. NOISE DETECTION AND RECOVERY PROCESS

Noise detection is based on the assumption that a noise-free image contains locally smoothly varying areas separated by edges. We assume the following noise model, any pixel at (i,j) in the noisy image can take on an arbitrary impulse noise n_{ij} from the dynamic range with probability equal to the noise rate p in the image. If the pixels $x_{ij,no}$ and $x_{ij,or}$ are in noisy and original images, respectively, then we have

$$x_{ij,no} = \begin{cases} x_{ij,or} & \text{with probability } 1\text{-}p \\ n_{ij} & \text{with probability } p \end{cases} \qquad (1)$$

A. Noise Detection

For the noise-free image, in the locally smooth region, the center pixel has numerous similar pixels among their neighbors in a filter window and the intensity difference is low. However, in the edge region, the pattern is very different. In the tangential direction of the edge, the intensity difference is the same as the smooth region, which we call as smooth similarity; whereas in the normal direction of the edge, the intensity difference is higher than the smooth similarity but not too high, which we call as edge similarity.

If the central pixel x and any pixel y of the other pixels in the window such that

$$abs(x-y) \leq T_{smooth} \qquad (2)$$

Then pixel x and pixel y are considered as smooth similarity, $Tsmooth$ is the threshold. If the pixel x and y satisfy

$$T_{smooth} < abs(x-y) \leq T_{edge} \qquad (3)$$

Then they are considered as edge similarity, $Tedge$ is the threshold.

For the noise-free or the noisy image, in the smooth region, the center pixel x of a filter window has many neighboring pixels that satisfy smooth similarity; whereas in the edge region, the center pixel x has little neighboring pixels that satisfy smooth similarity and some neighboring pixels that satisfy edge similarity. For the center pixel x, we denote Nss as the number of pixels that satisfy smooth similarity and Nes as the number of pixels that satisfy edge similarity in a filtering window. Now for a 3×3 window, the center pixel x is considered as the original pixel located in the smooth region if

$$N_{ss} > 3 \qquad (4)$$

And the center pixel x is reckoned as original pixel located in edge region if

$$N_{ss} == (2,3) \& N_{es} > 2 \qquad (5)$$

Other pixels are considered as noisy pixels.

B. Restoration Process

To restore the corrupted image, we replace each detected noisy pixel with a normalized weighted sum of its good neighboring pixels in the 3×3 filtering window. The weight is fixed and the weight of the original pixel in the smooth region is 4 and the weight of the original pixel in the edge region is 2. For those already restored noisy pixel, we also set a weight value which is depended on the intensity difference between the restored noisy pixel and the average intensity of its good neighboring pixels. If the intensity difference is small than 15, the weight is 4; If the difference is small than 25, the weight is 2; and in other cases, the weight is 1.

III. EXPERIMENTAL RESULT

To assess the performance of our proposed method, the standard gray-scale test images used in our experiments have distinctly different features. These images are "Lena" and "Walkbridge", each of size 512×512.

The peak signal-to-noise ratio (PSNR) between the restored image and the original image is selected as measuring the performance in experiments. The PSNR is defined as:

$$PSNR = 20\log_{10} 255 / MSE$$
$$MSE = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (y(i,j) - x(i,j))^2 \qquad (6)$$

Here MSE is the mean squared error and y denotes the deblurred image and x denotes the origin image.

A. Restoration Results

The noise density in the noisy images is varied from 20% to 60%. The thresholds that utilized in all the experiments are fixed with $Tsmooth$=15 and $Tedge$=50.

The proposed filtering algorithm is named as local similarity pattern (LSP). The PSNR resulting from various experiments is given in Table I for "Lena" image. From the table, it can be easily observed that the LSP outperform the other filtering schemes at noise levels below 40% and only SNC method outperforms our NSF method at higher noise levels.

TABLE I.   COMPARISON OF RESTORATION RESULTS IN PSNR (DB) FOR "LENA" IMAGE NOISE RATE

Method	20	30	40	50	60
Median	32.5	27.7	23.5	20.0	17.4
PSM[1]	27.8	26.5	25.4	23.6	20.6
MSM[2]	33.3	31.2	29.0	26.5	23.4
TSM[3]	31.5	27.5	24.1	21.5	19.6
EM[4]	34.8	32.3	30.0	27.6	25.6
PWMAD[5]	30.9	26.5	22.7	20.1	18.1
UNR[6]	33.3	31.2	28.6	26.1	23.7
SNC[7]	34.0	32.3	31.0	30.0	29.1
FSM[8]	26.3	22.2	19.2	16.6	14.4
LSP	34.9	32.6	30.0	27.4	24.2

Fig. 1 shows the restored images of our method at 20, 40, and 60% noise density for "Lena" image. It can be seen that the proposed method successfully preserve the details in the image while removing the noise at 20% noise rate, and have a scattered amount of tiny bright and dark spots at the 40% rate. It is clear that the restored image at 60% rate is seriously corrupted with a cluster of impulse noise. For convenience of visual judgment, the detailed regions cropped from Fig. 1 are shown in Fig. 2. From the Fig. 2, it shows the proposed algorithm can remove the noise successfully while preserving edges under 40% rate. Fig. 3 shows the restored images at 20, 40, and 60% noise density for "Walkbridge" image.

The reason is that many of the noisy pixels in the cluster are similar to each other but not similar to the other surrounding original pixels. That means the intensity differences between the noisy pixels are not too high and thus they are likely to be considered as original pixels located in the edge region when compared with the threshold $Tedge$ and $Tsmooth$ when heavily corrupted.
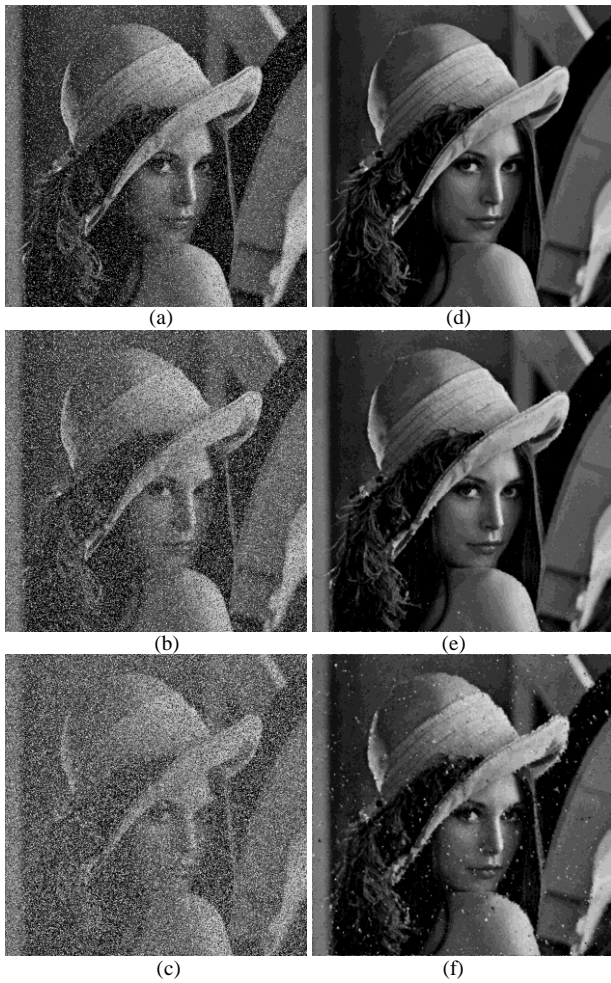
Figure 1.   Restoration performances of Lena image with different noise rate. (a)20% noise image(b) 40% noise image (c) 60% noise image (d) 20% denoise image (e) 40% denoise image (f) 60% denoise image
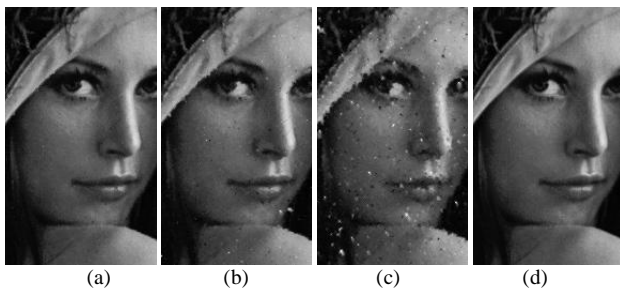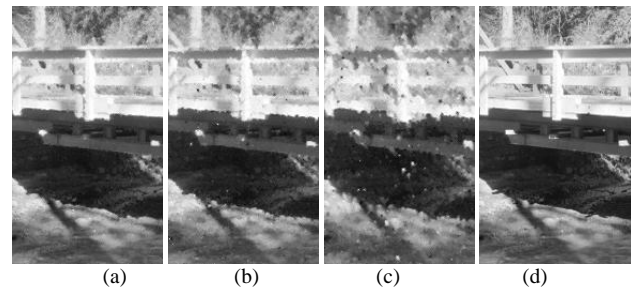

Figure 2.   Detailed regions corrupted from Fig. 1 (a) 20% denoise image (b) 40% denoise image (c) 60% denoise image (d) origin image

B. Noisy-Free Pixels Detection Results

The percentage of wrongly detected noise-free pixels (false detections), undetected noise-free pixels (miss detections) and rightly detected noise-free pixels (hit ratio) in the corrupted Lena and Walkbridge images are shown in Table II. It is clear that for the Lena image the proposed method has at least 90% hit ratio and at most 6% false rate or miss rate below 40% noise rate and the hit ratio decreases rapidly and false rate and miss rate increase rapidly when the noise rate is higher than 40%. This is why our method performs outstanding when noise rate is below 40% and performance decreases slowly with noise rate increasing. The detection results of Walkbridge image are

bad than Lena image for the miss and right ratio. Because the Walkbridge image has many edges and the edge region are hard to detect the noise pixel.
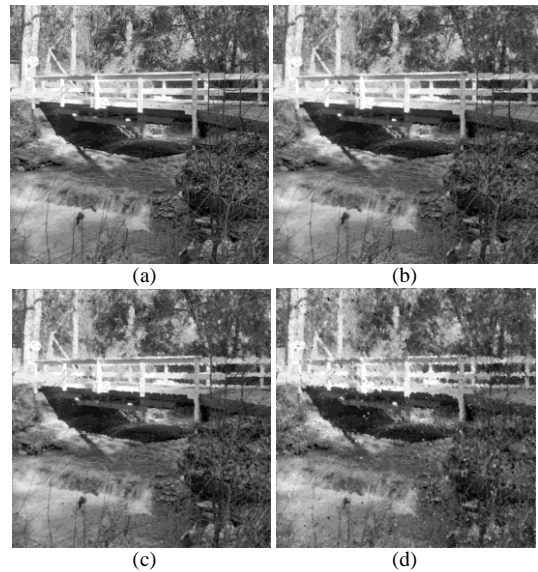
Figure 3.   Restoration performances of Walkbridge image with different noise rate. (a)20% (b) 40% (c) 60% (d) origin image

Figure 4.   Detailed regions corrupted from Fig. 3 (a) 20% denoise image (b) 40% denoise image (c) 60% denoise image (d) origin image

TABLE II.   NOISY-FREE PIXELS DETECTION RESULTS FOR "LENA" AND "WALKBRIDGE" IMAGE

Lena			
Noise rate	False	Miss	Right
10	1.7	1.3	97.9
20	3.7	2.6	96.6
30	6.0	6.1	93.1
40	8.8	13.4	85.8
50	11.5	25.2	74.1
60	15.0	40.7	58.5
Walkbridge			
Noise rate	False	Miss	Right
10	2.1	7.9	91.3
20	4.6	11.1	88.1
30	7.3	16.9	82.3
40	10.4	25.4	73.8
50	13.7	36.0	63.2
60	16.7	49.4	49.7

C. Parameters Robustness

For evaluating the robustness of the proposed thresholds, we calculate the values of PSNR for "Lena" and "Walkbridge" images which have distinctly unique features. "Lena" image has few fine details such as edge while "Walkbridge" image has many fine details.
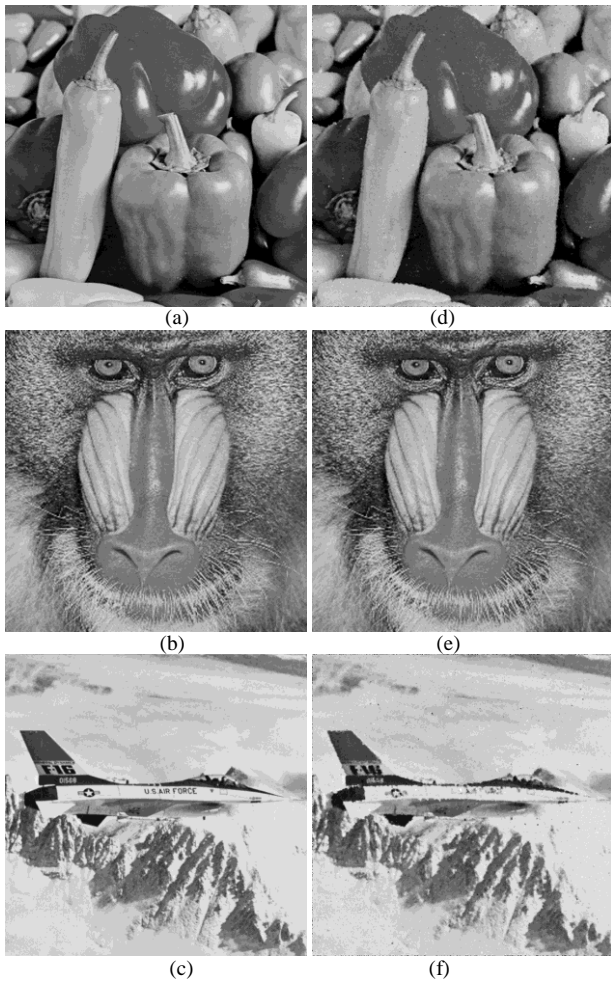
Figure 5.   Restoration for corrupted images with 40% noise rate (a) Peppers(b) Mandril (c) Jetplane (d) Peppers result (PSNR=30.15) (e) Mandril result (PSNR=25.71) (f) Jetplane result (PSNR=27.55)

First we fixed *Tedge* equal to 50, to study the *Tsmooth*'s impact on the performance, the results list in Table III. It can be shown that there is a gradual improvement in PSNR values as the threshold increase, until we reach the optimal threshold, then PSNR values decrease for both images. It is obvious that the *Tsmooth*'s role on restoration performance for the two images is distinct. Results of "Lena" are sensitive to *Tsmooth* whereas the results of "Walkbridge" are not. The optimal threshold is around 15 for "Lena" image whereas the optimal threshold varies greatly for "Walkbridge" image and greater than the Lena's. But fortunately, the restored results are good enough for the two types of images for any noise rate when *Tsmooth* in the [10, 20] range. Lastly, results of "Lena" image outperform results of "Walkbridge" image greatly.

For the small *Tsmooth,* the smooth similarity is harder to satisfied. This means that the hit ratio is low and there are many undetected noise-free pixels. So the restoration results are not good. While for large *Tsmooth,* the smooth similarity is easy to satisfied. This means that the false detection is high and there are many wrongly detected noise-free pixels which can form the dark or light spots in the restoration stage. So the results are not good as well. So the optimal *Tsmooth* is not very small and very large.

In the smooth regions, we need smaller *Tsmooth* to decrease the false detection while maintain high enough hit ratio, while in the edge regions, we need larger *Tsmooth* to increase the hit ratio while maintain small enough false detections. The Lena image has a few details and the majority regions are smooth regions, so the restoration results are major determined by the smooth regions. While the Walkbridge image has many details, so the restoration results are determined both by the smooth regions and edge regions. So the optimal *Tsmooth* of Lena image is smaller than that of Walkbridge image. It is harder to detect the noisy-free pixels in the detail region than the smooth region, so the restoration results of Lena is better than Walkbridge image.

Then we fixed *Tsmooth* equal to 15, to study the *Tedge*'s impact on the performance. The results list in Table IV. *Tedge*'s large scale change has little influence on the filtering results. The optimal threshold is around 50 for the two types of images for any noise rate.

Lastly, we test many other images and found that the *Tsmooth* and *Tedge*'s impact on filtering results are same as the "Lena" and "Walkbridge" images. This shows that the algorithm can achieve well enough filtering results for any image type at different noise density when *Tsmooth*=15 and *Tedge*=50. Fig. 5 shows the output images of assorted images for 40% noise density. It can be seen that the proposed method successfully preserve the details while removing the most noise and only scattered tiny noise spots left. Fig. 6 shows the output images of assorted images for 20% noise density.
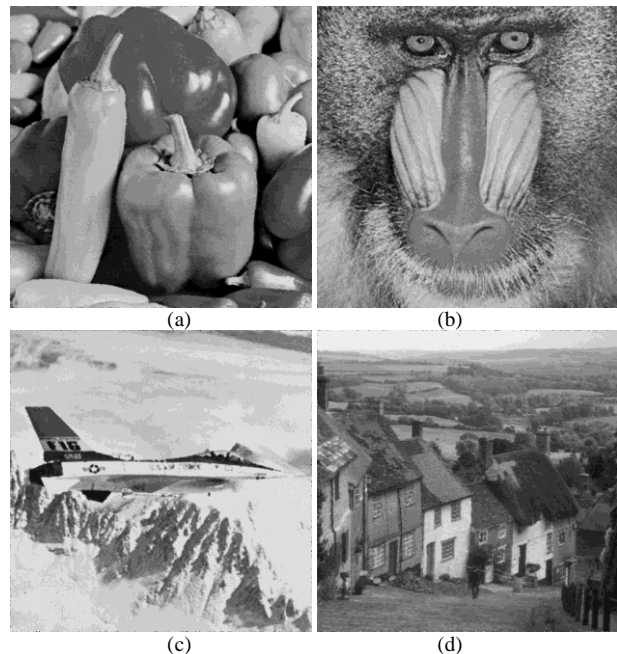

Figure 6.   Restoration for corrupted images with 20% noise rate (a) Peppers result (PSNR=33.33) (b) Mandril result (PSNR=28.47) (c) Jetplane result (PSNR=31.63) (d) Goldhill result (PSNR=32.14)

We also calculate the percentage of the pixels that have intensity difference in the filter window in Fig. 7. It shows that for all images, the curve have the nearly same profile. For Lena and peppers images, both the origin and 60% noise image, have the same curve shape. Before 10

TABLE III.   DEPENDENCE OF PSNR ON THE PARAMETER T_{SMOOTH} ON THE IMAGES (A) "LENA" AND (B)"WALKBRIDGE"

(a) Lena								
T_{smooth}	5	10	15	20	25	30	35	40
10%	32.83	35.96	36.94	37.18	36.82	35.99	35.16	34.32
20%	31.32	33.84	34.66	34.39	33.73	32.81	31.91	30.98
30%	29.78	31.83	32.50	32.14	31.40	30.49	29.49	28.63
40%	27.80	29.64	30.01	29.59	28.96	28.10	27.06	26.05
50%	26.19	27.50	27.54	26.96	26.21	25.32	24.26	23.21
60%	23.84	24.43	24.23	23.59	22.69	21.78	20.70	19.72
(b) Walkbridge								
T_{smooth}	5	10	15	20	25	30	35	40
10%	25.06	27.10	28.15	28.87	29.26	29.46	29.56	29.71
20%	24.22	25.95	26.78	27.19	27.33	27.34	27.31	27.26
30%	23.08	24.65	25.38	25.64	25.70	25.67	25.53	25.40
40%	22.06	23.55	24.11	24.24	24.18	24.10	23.93	23.70
50%	20.99	22.15	22.60	22.69	22.64	22.46	22.13	21.76
60%	19.72	20.70	20.93	21.04	20.89	20.65	20.27	19.82

TABLE IV.   DEPENDENCE OF PSNR ON THE PARAMETER T_{EDGE} ON THE IMAGES (A) "LENA" AND (B) "WALKBRIDGE"

(a) Lena								
T_{edge}	20	35	50	65	80	95	110	125
10%	34.19	36.09	37.70	38.09	38.10	38.19	38.13	37.90
20%	31.88	33.18	34.54	34.92	34.72	34.43	34.06	33.59
30%	30.26	31.47	32.20	32.05	31.61	31.06	30.42	29.75
40%	28.55	29.69	30.13	29.71	28.77	27.85	27.11	26.44
50%	26.54	27.16	27.18	26.37	25.45	24.56	23.70	23.06
60%	24.50	24.65	24.09	23.06	21.99	21.16	20.50	19.98
(b) Walkbridge								
T_{edge}	20	35	50	65	80	95	110	125
10%	25.62	27.08	28.26	28.81	29.02	29.11	29.13	29.13
20%	24.58	25.77	26.71	27.16	27.28	27.28	27.23	27.12
30%	23.58	24.62	25.46	25.77	25.82	25.72	25.55	25.35
40%	22.43	23.38	24.06	24.30	24.23	23.98	23.70	23.38
50%	21.24	22.00	22.61	22.67	22.47	22.15	21.80	21.47
60%	19.94	20.57	20.99	20.86	20.47	19.97	19.57	19.18

intensity difference, the curves decline significantly and after 20 intensity difference, the curves remain constant. So this is why Lena filtering results are sensitive to *Tsmooth* and the optimal value is between 10 and 20. For Walkbridge and mandril images, they also have the same curve shape, but the shape is different from Lena's. The curve decline gradually after 10 intensity difference. So this is why walkbridge filtering results are not sensitive to *Tsmooth* and the optimal value is between 15 and 35.

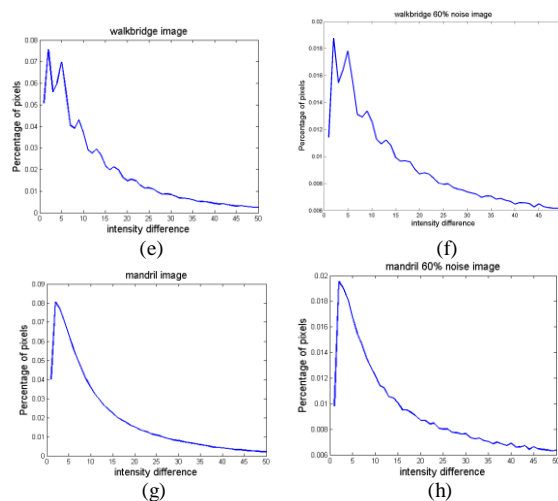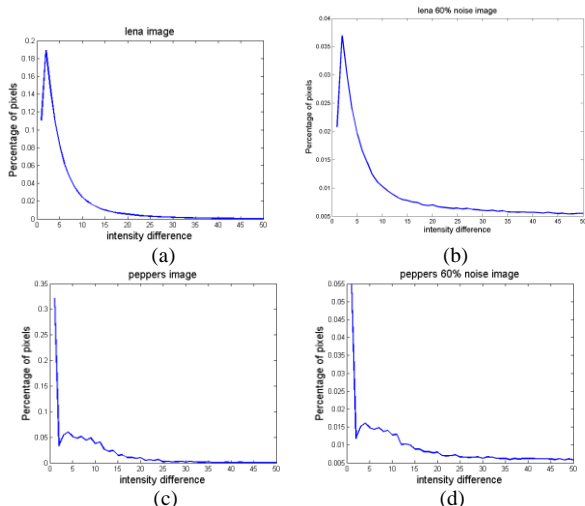Figure 7.   The percentage of the pixels of intensity difference with different noise rate. (a)Lena origin image(b) Lena 60% noise image (c) peppers origin image (d) peppers 60% noise image.(e) walkbridge origin image (f) walkbridge 60% noise image.(g) mandril origin image (h) mandril 60% noise image

D. Computational Complexity

Our method only needs to calculate the similarity numbers which just need to calculate the eight absolute intensity differences and at most 16 comparison operators for each pixel only once for the impulse detection. In the restoration process, we use fixed weight average method

to restore the noisy pixel which in 3×3 filtering window. Our filter only use integer type which is faster than float type. Whereas most recently proposed filters are performed in an iterative manner (the number of iterations is typically around 5) and need time-consume sort operator and varying filter window size (the window size of 3×3 to 7×7), and use much slowly float operation. So our method is faster than most of the methods available in the literature. For example, the runtime of our method is about 40ms for 512×512 gray-scale images written in c language in the 2.4GHz CPU platform.

IV. CONCLUSION

In this letter, a new noise removal method is proposed in this paper based on local similar pattern. For any pixel to be considered as an original pixel, it should have sufficient number of similar pixels amongst its neighboring pixels in the local window. The efficacy of the proposed method is evidenced by extensive simulations. The experimental results exhibit significant improvement in performance over several other methods. Furthermore, the proposed method is very straightforward and faster.

ACKNOWLEDGMENT

REFERENCES

[1] Wang Z, Zhang D, Progressive switching median filter for the removal of impulse noise from highly corrupted images, *IEEE Transactions on Circuits and Systems II*, vol. 46, no. 1, pp. 78-80, Jan. 1999.

[2] Chen T, Wu HR, Space variant median filters for the restoration of impulse noise corrupted images, *IEEE Transactions on Circuits and Systems II,* vol. 48, no. 8, pp. 784-789, Aug. 2001.

[3] Chen T, Ma KK, Chen LH, Tri-state median filter for image denoising, *IEEE Transactions on Image Processing,* vol. 8, no. 12, pp. 1834-1838, Dec. 1999.

[4] Luo W, Dang D, An efficient method for the removal of impulse noise, *IEEE international conference on image processing*, Atlanta, 2006, pp. 2601-2604

[5] Crnojevic V, Senk V, Trpovski Z, Advanced impulse detection based on pixel-wise MAD, *IEEE Signal Processing Letters,* vol. 11, no. 7, pp. 589-592, Jul. 2004.

[6] Garnett R, Huegerich T, Chui C, et al. A universal noise removal algorithm with an impulse detector, *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1747-1754, Nov. 2005.

[7] Ali Said Awad, Hong Man, Similar neighbor criterion for impulse noise removal in images, *Int. J. Electron. Commun. (AEU)*, vol. 64, no. 5, pp. 904-915, May. 2010.

[8] S. Zhang, M A Karim, A new impulse detector for switching median filters, *IEEE Signal Process. Lett*, vol. 9, no. 11, pp. 360-363, Nov. 2002.

[9] Shoushui Chen, Xin Yang, G. Cao. Impulse noise suppression with an augmentation of ordered difference noise detector and an adaptive variational method, *Pattern Recognition Letters,* 2009, 30 pp. 460-467.

[10] Garnett R, Huegerich T, Chui C, et al. A universal noise removal algorithm with an impulse detector, *IEEE Transactions on Image Processing* 2005, 14(11) pp. 1747-1754.

[11] Xuming Zhang, Youlun Xiong, Impulse Noise Removal Using Directional Difference Based Noise Detector and Adaptive Weighted Mean Filter, *IEEE Signal Processing Letters,* Vol. 16, No. 4, April 2009.

[12] Joan-Gerard Camarena, Valent ń Gregori, Samuel Morillas, Almanzor Sapena, Two-step fuzzy logic-based method for impulse noise detection in colour images, *Pattern Recognition Letters.* 2010 31 pp. 1842-1849.

[13] Hsiang-Chieh Chen, Wen-June Wang. Efficient impulse noise reduction via local directional gradients and fuzzy logic, *Fuzzy Sets and Systems.* 2009 160 pp. 1841-1857

[14] Fabijanska A, Sankowski D, Noise adaptive switching median-based filter forimpulse noise removal from extremely corrupted images, *IET Image Process,* 2011;5(5) pp. 472-80.

[15] Esakkirajan S, Veerakumar T, Subramanyam AN, Chand CHP. Removal of high density salt and pepper noise through modified decision based unsymmetric trimmed median filter. *IEEE Signal Process Lett* 2011;18(5) pp. 287-90.

[16] Tripathi AK, Ghanekar U, Mukhopadhyay S, Switching median filter: advanced boundary discriminative noise detection algorithm. *IET Image Process,* 2011, 5(7) pp. 598-610.

[17] S. Akkoul, R. Lédée, R. Leconge, R. Harba, A new adaptive switching median filter, *IEEE Signal Process. Lett.* 17 (2010) pp. 587-590.

[18] Zhiyong Zuo, Tianxu Zhang, Jing Hu, Gang Zhou, A new method for removing impulse noise based on noise space characteristic, *Optik* 124 (2013) pp. 3503-3509

[19] Chai Ling, Adaptive Image De-noising Algorithm in Intersecting Cortical Model, Journal Of Multimedia, Vol. 8, No. 5, October (2013) pp. 534-540

[20] Gang Liu, Jing Liu, Quan Wang, Wenjuan He, The Translation Invariant Wavelet-based Contourlet Transform for Image Denoising, *Journal Of Multimedia,* Vol. 7, No. 3, June(2012) pp. 254-261A

[21] Pan Duan, Kaigui Xie, Na Song, Qichang Duan, Method of Vehicle License Plate De-noising and Location in Low Light Level, *Journal Of Networks,* Vol. 5, No. 12, December (2010) pp. 1393-1400

Shan Jianhua received the B.S. and Ph.D Degree in precision mechanism from University of Science and Technology of China, China, in 2002 and 2007, respectively. From 2007 to now, he worked as a Assistant Professor in Anhui University of Technology. His research interests include image processing and robot.

Zhu Liangliang received his bachelor in 2012 from the Department of Mathematics and Applied Mathematics in Anhui University of Technology. Currently, he is studying in the Department of Mechanical Engineering in Anhui University of Technology for the M.S. Degree. His main research directions is image processing.

The Rules of Attention Shift on Display and Control Terminal Base on Situation Awareness

Liu Wei [1, 2] and Zhang Bo [1, 2*]

1. Beijing University of Posts and Telecommunications, Beijing, China
2. Beijing Key Laboratory of Network Systems and Network Culture, Beijing, China
*Corresponding author, Email: twhlw@163.com, zb2242@gmail.com

Abstract—**To investigate the rules of a user's attention allocation and attention shift when they are using a display control terminal, using both methods of information processing, situation awareness theory and eye-movement tracking technology were applied to analyze the influence of different interface element layouts on attention allocation and attention shift. In this study, 26 participants performed an operating task under different display control interfaces, which were divided into two types and used to simulate different situation awareness levels, and the fixation point distribution was recorded as the evaluation index. The participants were asked to perform a daily washing task and a fixed task on the simulation interface of a cylinder washing machine. The experimental results revealed that different situations and different situational awareness levels of the participants influenced the rule of attention allocation and transfer; the information elements on the interface play a key role in the user's cognitive process. The experimental results and users' subjectivities were generally in agreement; thus, the present study could provide ergonomic evidence with display and control terminal interface design.**

Index Terms—**Situation Awareness; Attention Shift; Eye-Movement Tracking; Cognitive Engineering; Interface Design**

I. INTRODUCTION

A display and control terminal in accordance with a harmonious relationship between a human, a computer and the environment is designed based on the efficient acquisition of visual information and rational attention distribution. Scientifically, human-computer interfaces are developing into complicated and diverse entities; thus, display control interface usability research has reached great importance in the accurate comprehension of visual scanning behavior and the gathering of efficient information about attention distribution. Attention, when processing the information, is usually regarded as an important psychological adjustment mechanism, which can be attributed to the limited information-processing resource and highlight the selective function of perception [1]. There are several studies on the modeling of attention distribution. Senders *et al.* established the first quantitative model of instrument monitoring behavior to analyze visual scanning behavior in the context of the intelligence bandwidth concept [2]. The attention model established by Itti *et al.* combines

different visual features such as brightness, color and orientation [3]. Wickens translates the factors affecting attention distribution into information prominence, diligence, expectation and value, and he set up an SEEV model that considers these factors [4]. Nobuyuki *et al.* were the first to establish the attention distributive model of car drivers and to explain the relationship between psychological thinking activity and attention model by using a fuzzy control model [5]. Wu Xu and his colleagues from the Beijing University of Aeronautics and Astronautics started using two main sources of information and finally set up an attention distribution prediction model based on the significance, probability, prominence and diligence of information [6]. However, these models are all based on the dominant character of interface elements, and they predict the attention distribution with the subjective expectation by the users who have certain intentions when using a given interface.

Opposite to these previous results, attention shifts during the task; therefore, it is necessary to realize the dynamic distribution in accordance with different cases. Usage scenarios and using experiences have often been considered to be a major source of user distraction potentially diverting attention away from one functional area to another. Some operational error reviews have even related different situations and different situational awareness levels of the participants influenced attention allocation and transfer even under a same interface. Unfortunately, prior ergonomics investigations have not sufficiently addressed the eye fixation and fixation time distribution related to situation awareness.

In visual activities, attention is highly related to eye movement. In general, eye movement can best show the selective function of attention in visual information processing. The results from several studies prove that attention transformation is highly associated with eye movement during many visual tasks [7, 8, 9, 10]. Moreover, the position of attention transformation and eye movement is the same. By analyzing the general situation of users using the display and control interface of the roller washing machine with classical tasks, this study focuses on analyzing the transformative statement of attention in different functional areas under different tasks. Comparisons are made among the four interfaces to give advice on the arrangement of display and control interface.

In the present study, we sought to complement prior research by specifically addressing situation awareness that might effects on attention distribution and transition setting by using four high-fidelity roller washing machine simulation interfaces (Fig. 1 shows the interfaces tested in the study). The objective was to compare participants perceptual and performance responses with these four interfaces, which have different designs of five functional areas: power, start, program, button, and screen. The utility of the present study is how the dominant character of interface elements and different situation awareness levels may contribute to participants' attention and transition distraction based on roller washing machine interface and performance degradations.

Based on the literature review [11, 12, 13], it is possible that adjacent elements, and/or prominent elements might lead to greater participant eye fixation time and an increased potential for distraction. Our primary hypothesis was that different designs of functional areas would lead to different eye fixation distribution and operational performance. It was also expected that different situation awareness levels might lead to attention diversion in the same interfaces.

II. SITUATIONAL-AWARENESS-ORIENTED DIVISIONS OF USER MODELS

According to the SA model by Endsley [14, 15, 16], to perceive the current situation, to analyze and comprehend the current situation as well as to predict the future situation, we divided our users into three types.

(1) Inexperienced users. These users do not know anything about the washing machine interface. In daily life, they are few in number and are likely to be young people who have recently begun living independently or they could be children. Without any experience and previously developed habits, they use the interface with assistance from the instructions or they use a bottom-up attempt.

(2) Inexperienced users. These users know something about using the control interface of the electric appliances, including new appliances. With some common sense, such users usually try to use the interface from their limited experience, and have a top-down attempt to adapt to the new interface, yet they still have to make a bottom-up attempt again to master certain new functions.

(3) Experienced users. They have a well-organized experience in dealing with interfaces. Normally, they have possessed electric appliances for a long time, or they are senior citizens who are unwilling to change their life style. With outstanding experience, such users will usually try the top-down method.

III. METHODOLOGY

A. Apparatus

To study the real situation of interface-using, we made electronic changes in four interfaces of roller washing machines by creating four flash models as the experimental and comparison objects (see Fig. 1).

The experimental interfaces were put on the Stimulus display end, with a resolution ratio of 1440×900, an average brightness of $120cd/m^2$, and an average illumination of 600lx. The SMI-RED from Germany was adopted to record the eye movement because it can trace the eye movement by an IR camera on the bottom, and the subjects can freely move their heads to some extent without any equipment. The SMI-RED also maintains accurate measurements.
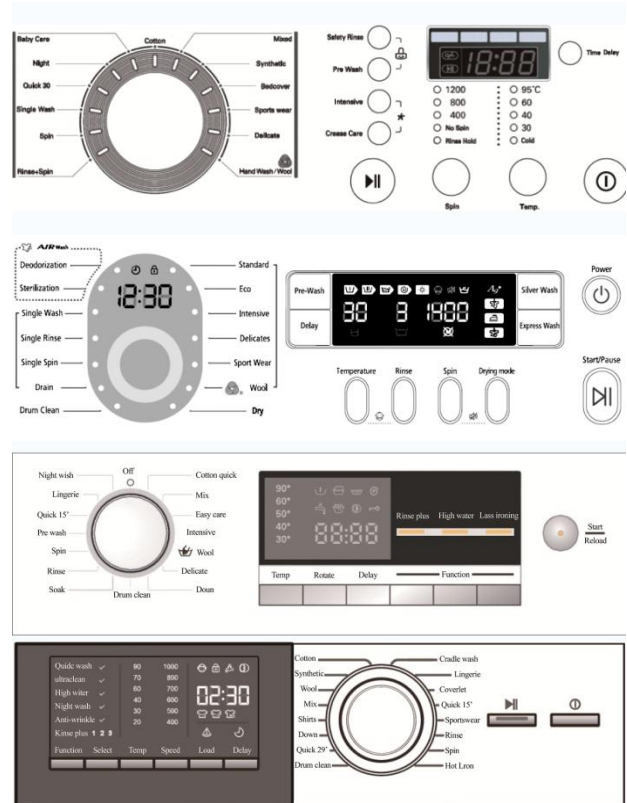

Figure 1.   Flash model of four machine interfaces

B. Participants

This experiment entailed interviewing 26 people, 20 of whom were 26- to 40-year-old females who frequently use the roller washing machine, have and use the roller washing machine at home, were right-handed, had visual acuity (including corrected visual acuity) over 1.0, and did not have color feebleness or color blindness. They are divided into 2 groups. Group A consisted of the inexperienced users who executed tasks without getting accustomed to the testing interfaces, while group B, the experienced people, were familiarized with the interfaces in advance.

C. Experiment Design

According to the general habits of the users, the testing interfaces were arranged differently and contained the following five elements: power buttons, start/pause, rotary knob, display screen and functional button as elements 1 to 5. The experiment maintains two parts: task and free use. The users got acquainted with the procedure of attention transformation features of different interfaces, and they talked about their overall feelings. Eye movement was traced in the whole experiment.

D. Experiment Procedure

(1) Warming up: introduce the method of using flash and tips of using eye tracker to the users and lead them to be seated at the chair 50 cm away from the display screen and adapt themselves into comfortable gestures. Calibration should be made on heads and eyes. Before each testing, the eye trackers should be controlled by nine matrices. The users were required to keep their head and body as stable as possible.

(2) Practice: before testing, the users will practice once on the testing interface.

(3) Testing: four interfaces will be tested individually, and for each test, control measurements should be conducted to ensure the accuracy of eye trackers.

(4) Two Tasks: there are continuous single and open tasks for each interface. Group A conducts a single task first and then open tasks, while group B is the opposite of group A. After each task, the users will be asked to mark on the interface, ranging from 1 to 5 points. The appearance of interfaces are random, so different users will finish the interfaces in a different order.

IV.   RESULT AND DISCUSSION

To know the attention distribution and scanning features of the users, we made a contrast between the inexperienced and experienced users to conclude how the elements affect attention. We also divided the five elements of the interface into 5 areas of interest (AOI, see Fig. 2). We analyzed the rational features of interface arrangement in terms of attention time, percentage of focus as well as matrixes of focus in every AOI.
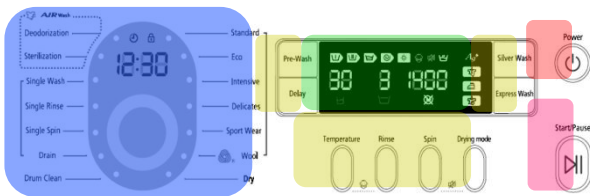
Figure 2.   Divide the interface into 5 AOI (interface 2)

A. How the Position, Size and Shape of Power Buttons Affect Attention

The task of turning on the washing machine is very easy. The inexperienced users and experienced users took a similar amount of time to complete this task except interface 3. We observed how the users make decisions by their scanning trace.

Interface 1: when attempting to locate the power button, users tended to look at the top right corner, and they finally found that it is in the bottom right corner. The overall tendency can be described in the picture below. All of the users finished this task swiftly and easily.

Interface 2: The inexperienced users first look at the top right corner and are attracted by the start button, thus making an attention transformation. They find it is not the target and return soon (see Fig. 4). They are affected by the start button, but they can still position it as quickly as the experienced users.
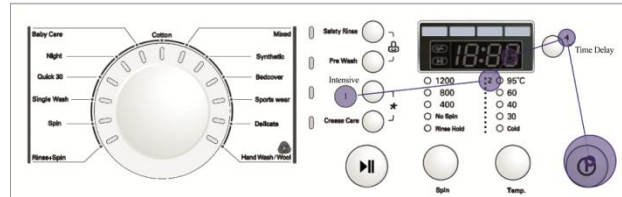

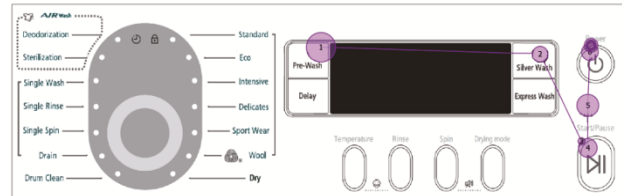Figure 3.   Saccades path of A05 by completing power task (interface1)

Figure 4.   Saccades path of A01 by completing the power task (interface2)

Interface 3: all of the inexperienced users have the same problem in this section: they are attracted by the start button with focus on the right side which is not right at all. They even touch the start button until the failure shows that they are wrong. Some of them cannot complete the task until the host gives them some information (see Fig. 5). Some experienced users make some mistakes too, but all of them can eventually complete the task (see Fig. 6). Inexperienced users complete this task slowly with great difficulty, some even failed to do it, while the experienced users performed much better.
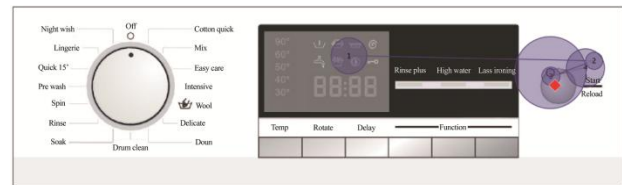

Figure 5.   Saccades path of A05 by completing the power task (interface3)
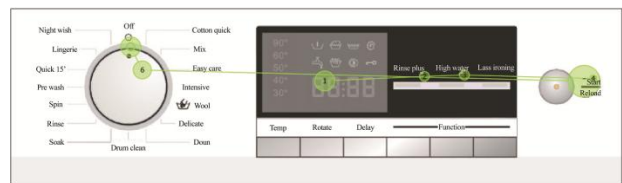

Figure 6.   Saccades path of B04 by completing the power tasks (interface3)

Interface 4: all of the users scan to the right side. Because the power button and start button have similar sizes, shape and color, with pictures but no words, the inexperienced users spend a lot of time considering the pictures (see Fig. 7), while the experienced users will quickly recognize the power button. The users need attention distribution to click correctly (see Fig. 8). Therefore, they cannot finish the task swiftly, especially when they have difficulties in positioning. The inexperienced and the experienced users do not differ in time, but rather in efficiency.
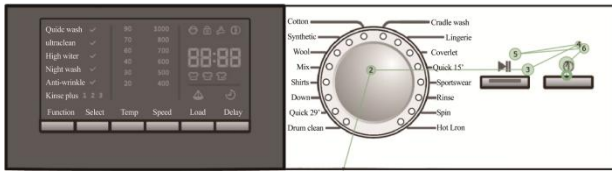
Figure 7.   Saccades path of A04 by completing the power task (interface4)
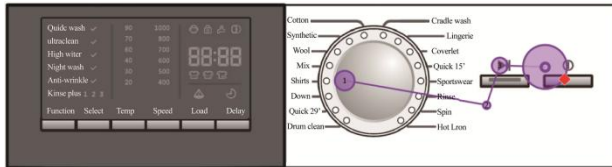

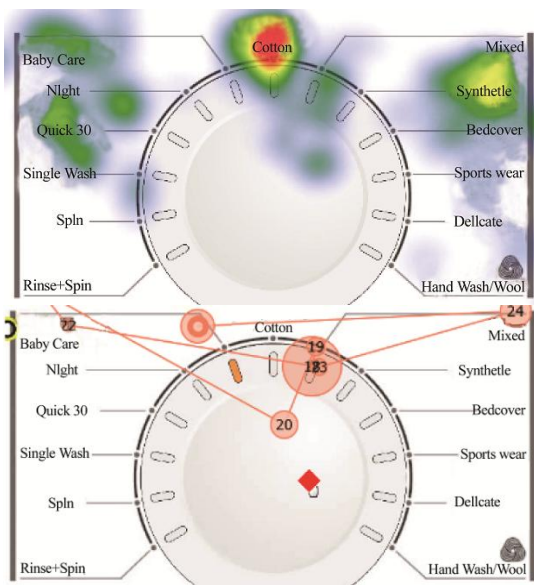Figure 8.   Saccades path of B01 by completing the power task (interface4)

Figure 9.   Heat map and Saccades path of the program knob (interface1)

It is well known that most human beings are right-handers. Generally speaking, inexperienced users will resort to their experience to have a top-down attempt, and they are likely to look the right side when they come across a new interface, easily being attracted by shape, size and color, when their experience cannot help (see interface 3). The attention distribution becomes urgent; they need to try bottom-up to become familiar with all of the elements, and when the arrangement of the interface is too creative, the users may be confused and cannot finish the task independently. The attention transformation of the experienced users is clear, with correct positioning, but when the shape of the button is dramatically changed (see interface 4), they may need more time to think.

B. How the Program Knob Distribution of the Washing Machine Affects Attention

Research shows that users care about the washing procedure as well as the different preset parameters; therefore, the users should choose their optimal procedure and their attention will remain in the area for a long time.

A well-organized arrangement can accelerate the process of selecting information. A heat map is used to analyze important information, and the warm color is used to mark dynamic changes in time and position.

Interface 1: (see Fig. 9) Users attach importance to the central part of the rotary knob because of their interest in it. The users are accustomed to look from the left to the right to find their desired procedure. Attention is focused more in the upper region, and the inexperienced users can position and swiftly.

Interface 2: Dotted box and brackets are used to divide this section, which makes it is very convenient for the users to find. Attention is focused on the procedure that they use most frequently. First, they will try an up and down method of scanning and they will stop at each part to judge whether it is their target. The inexperienced users can position easily and swiftly.
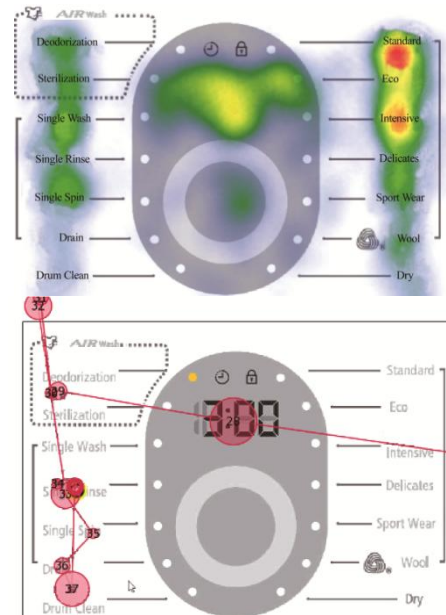

Figure 10. Heat map and Saccades path of the program knob (interface2)

Interface 3: The power function is combined with the rotary knob, and the power-control part is above, which is an area that the users are likely to look at. In the upper left-hand corner, the rotary knob is always noticed first by the users, but they seldom use it. The procedures that are often used are arranged to the right side. The users worry that when they control it from the left to the right side, they might touch the power button, so they tend to be more careful, which consequently affects the attention distribution transformation. The inexperienced users have difficulties in positioning, which makes them inefficient at completing this task.

Interface 4: in this interface, the types of procedures are optional. Different types of procedures are not arranged in a well-organized manner around the knob. The users need to read the instructions to find the procedures that they need. For example, there are two fast-washing procedures but they are arranged awkwardly such that the users cannot use it efficiently. It is very

difficult for users, especially the inexperienced users, to use this interface.
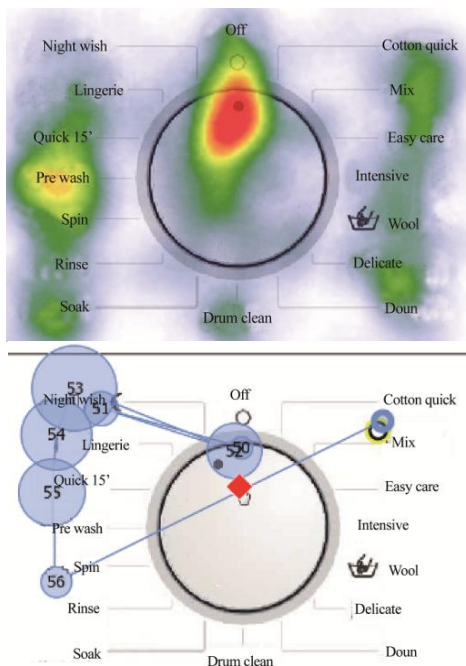

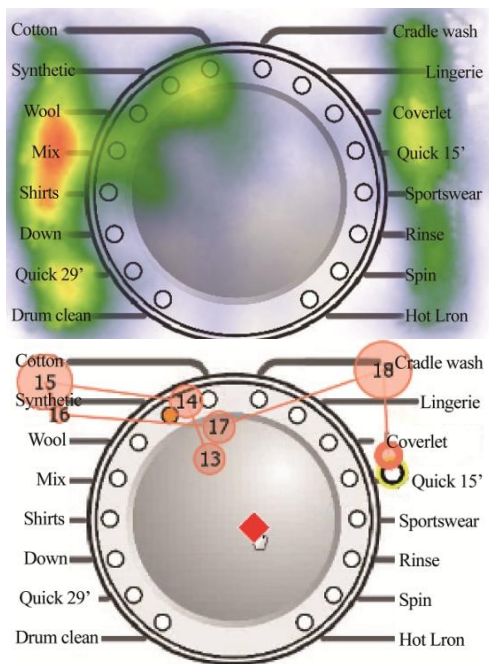Figure 11. Heat map and Saccades path of the program knob (interface3)

Figure 12. Heat map and Saccades path of the program knob (interface4)

In summary, the inexperienced users are accustomed to seek information in a left-to-right manner, judging from the overall classification and positioning of their targets, and then they find the targets in an up-down manner. Based on this analysis, the users' attention is mostly focused on the upper region. Therefore, it is useful to determine the procedures the users like to use, and arrange those in the upper areas.

C. How the Arrangement of Functional Buttons Affects Attention

The functional buttons are all arranged in a unique manner among the four interfaces to study how different arrangements affect attention. In this section, the eye movement between inexperienced users and experienced users is quite different, which suggests that the arrangement does affect the users greatly.

Interface 1: in this section, the users are required to start two functions of crease care followed by a 3-hour delay (see Fig. 13). When starting the former task, the inexperienced and the experienced users are similar; however, while starting the latter task, users have much more eye movement, especially the inexperienced users who require two times as much eye movement as the experienced users.
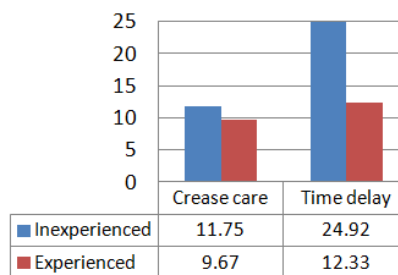

Figure 13. The average saccadic number cost to complete open special functions task by novice and experienced users

In this interface, with the crease care functional button being near the knob, users can easily position it. However, the time delay button is too far away from other buttons, which is easily ignored by users. Moreover, there is another button that is very similar to this one, which may mislead the users. According to the habit of the users (top-down), all of the functional buttons should be located in close proximity to each other. If one functional button is too far away from its group, the users may have difficulty finding it, they might find it with lower efficiency, or they might even fail to find it unless the host assists them. Fig. 14 and Fig. 15 are the scanning traces of inexperienced users.
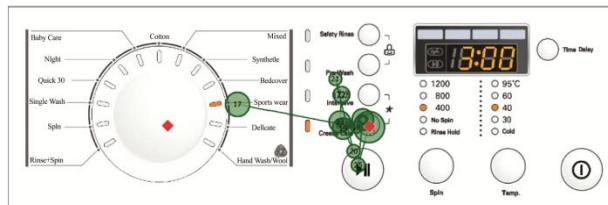

Figure 14. Saccades path of A03 by completing the open crease care task
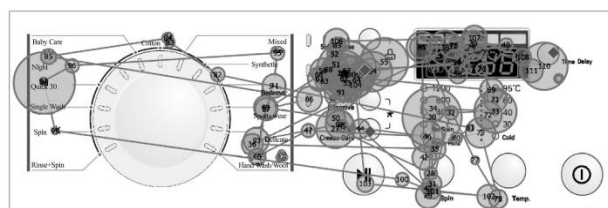

Figure 15. Saccades path of A03 by completing the time delay task

Interface 2: in this interface, the users are required to start the functional button of "silver wash." When they try it, they read the functional buttons on the left first, and then read the parameter button below the screen, and then look from left to right. After a series of eye movement, they finally complete the task. In this interface, the four functional buttons are arranged on the two sides of the screen, and the first side near the knob can be easily noticed by the users, while the other is much more difficult to be noticed. The scanning trace of inexperienced users is described in Fig. 16.
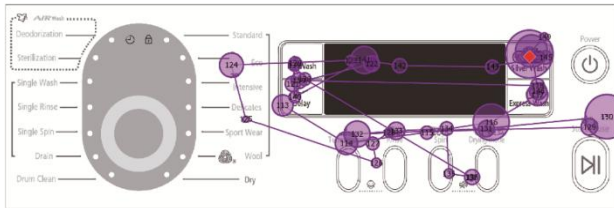

Figure 16. Saccades path of A02 by completing the open silver wash task

Interface 3: in this interface, the users are required to set the rotation speed to 1200, and they also need to start the functional button of the high water level. Most of the inexperienced users mistake the written instructions above the button for the button. Through the eye movement, we observe that they do not notice the button, but they just click the words instruction directly. In our interview, they reply that they just regarded the words as the button, and the button as a dent in the screen.
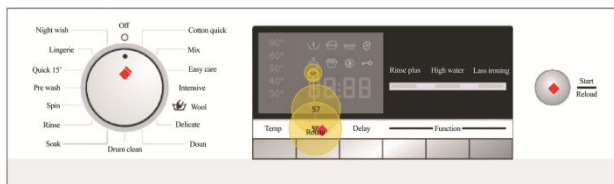

Figure 17. Saccades path of A11 by completing the adjust speed task

The high water level functional button is separated from the area of the functional buttons with vague written instructions so inexperienced users cannot comprehend their relationship clearly, and they mistook the words as the buttons, which in their eyes is not user-friendly. In such way, users cannot use it efficiently.
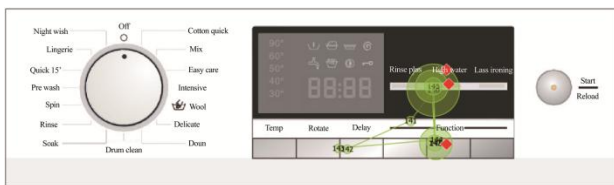

Figure 18. Saccades path of A12 by completing the high water level task

Interface 4: in this interface, all of the functional buttons are in the functional menu. The users can freely choose it by clicking different functional buttons, and they can click the optional button to start. Although all of the users can use it correctly, most of them do not like it.

They prefer a design of separated functional buttons. In interface 4, the users need to click the menu button several times, and their eye movement correlates with the cursor, which costs a lot of time and results in lower the efficiency. Experienced users do not think highly of this arrangement; they prefer separated buttons which are easier to locate.

The functional button design of four interfaces shows that the inexperienced users are likely to search for the target among adjacent elements; thus, the buttons with similar function can be located together near the knob or the screen. The buttons and their written instructions should be well-organized which can enable the users to locate them efficiently. Each functional button should have its independent button as much as possible, so that the users will not need to read the instruction for help. When the users are accustomed to a certain functional button, they may find it directly and swiftly. This is especially true for the experienced users.

D. The Rules of Attention Transition Among AOI

By tracing the eye movement in the four interfaces, we calculated the probability of attention transition from one area to another, which enabled us to determine the general rules of attention transition among different AOI.

From Fig. 19 and 20, the transition of attention in interface 1 and 2 is not logically well-organized. The users can be easily attracted by fresh elements or adjacent elements. Their eye movement does not follow a general pattern, and their efficiency of use is poor.
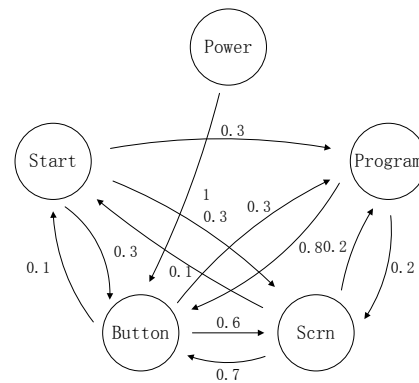

Figure 19. Transition probability of the user's attention among AOIs (interface1)
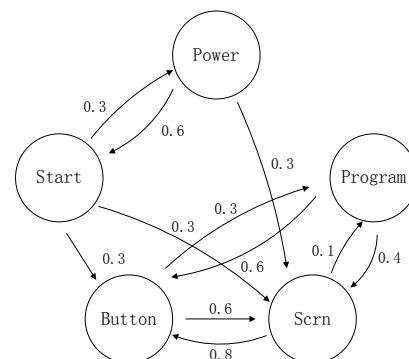

Figure 20. Transition probability of the user's attention among AOIs(interface2)

As shown in Fig. 21 and 22, the users of interface 3 and 4 can basically form a transition process in the following order: power button, washing procedure, display screen, functional buttons and start button. This process is in accordance with the usage habits.
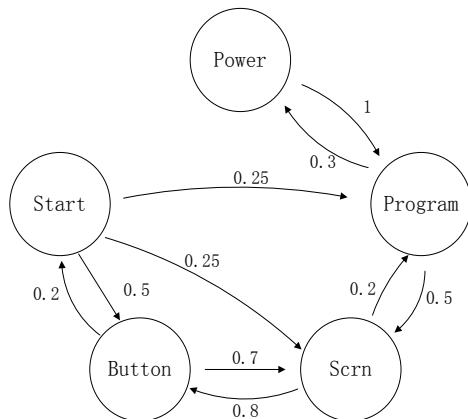

Figure 21. Transition probability of the user's attention among AOIs (interface3)
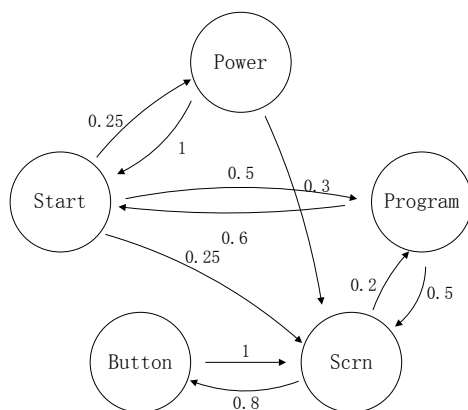

Figure 22. Transition probability of the user's attention among AOIs (interface4)

V. CONCLUSIONS

(1) Different situations have different effects on attention distribution and transition. The inexperienced users need to process the element information, which is in top-to-bottom order [17]. According to this research, we have knowledge of how the interface elements affect attention in different situations.

(2) Under the circumstance of different levels of consciousness, people can perceive the significance and probability of the information to some extent. By processing the information features, the attention can shape a working memory, which is from top to bottom and in accordance with the feature-integration theory of attention [18]. The eye movement data provided by this study is in accordance with the users' subjectivity, which strengthens our conclusion.

A visual display terminal should be in accordance with a harmonious relationship between humans, computers and the environment. Its design should be based on the efficient acquisition of visual information and rational attention distribution. Only by combining these two effects scientifically can we design products with a harmonious relationship between humans, computers and the environment.

REFERENCES

[1] Ding J Z, Zhang Q, Guo C Y, et al. Cognitive psychology. *China Remin University Press,* 2010 pp. 41

[2] Senders J W. The human operator as a monitor and controller of multidegree of freedom systems. *Human Factors in Electronics,* 1964, 5(1) pp. 2

[3] Itti L, Koch C. Computational modeling of visual attention. *Nature Reviews Neuroscience,* 2001, 2(3) pp. 194

[4] Wickens C D, Alexander A L. Attentional tunneling and task management in synthetic vision displays. *The International Journal of Aviation Psychology,* 2009, 19(2) pp. 182

[5] Nobuyuki Matsui, Eiichi Bamba. Consideration of the attention allocation problem on the basis of fuzzy entropy. *Association Symposium of Measurement and Automatic Control,* 1996, 22(12) pp. 27

[6] Wu X, Wanyan X R, Zhuang D M. Attention allocation modeling under multifactor condition. *Journal of Beijing University of Aeronautics and Astronautics.* 2013, 8(39) pp. 1086

[7] Yu Z, Elizabeth H, Meghan R, et al. Driver distraction and performance effects of highway logo sign design. *Applied Ergonomics,* 2013, 44 pp. 472

[8] Liu W, Yuan X G. Comprehensive evaluation on pilots situation cognitive. *Space Medicine & Medical Engineering.* 2008, 36(02) pp. 168

[9] Hiroyuki S, Duk S, Takeshi K, et al. Attentional effects on gaze preference for salient loci in traffic scenes. *Ergonomics.* 2012, 7 pp. 743

[10] Wilson M, Chattington M, Marple-Horvat D E, et al. Eye movements drive steering: Reduced eye movement distribution impairs steering and driving performance. *Journal of motor behavior.* 2008, 40(3) pp. 190

[11] Prasanna R, Azizah J, Farizah H A R, et al. Evaluation of user Interface Design for Leaning Management System (LMS): Investigating Student's Eye Tracking Pattern and Experiences. *Social and Behavioral Sciences.* 2012, 12 pp. 527

[12] Martijn S, Angela R, Monica R, et al. User-centered design in brain–computer interfaces—A case study. *Artificial Intelligence in Medicine.* 2013. 10 pp. 71

[13] Ana C B, Roxanne L, Jorge R. Design and Evaluation of a Mobile User Interface for Older Adults: Navigation, Interaction and Visual Design Recommendations. *Procedia Computer Science.* 2014 pp. 369

[14] Endsley M R, Bolstad C A. Individual differences in pilot situation awareness. *International Journal of Aviation Psychology.* 1994, 4 pp. 241

[15] Endsley M R. Towards a theory of situation awareness in dynamic systems. *Human Factors: Situation Awareness Special Issue.* 1995, 37 pp. 32

[16] Endsley M R. Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society 32nd Annual Meeting.* 1988, Anaheim

[17] Treisman A, Gelade G A. A feature integration theory of attention. *Cognitive Psychology,* 1980, 12(1) pp. 97

[18] Kim M S, Cave R C. Perceptual group via spatial selection in a focused attention task. *Vision Research,* 2001, 41(5) pp. 611

Liu Wei, 1970.6, Shandong, Chinese, Male, Professor, Doctor, HCI cognitive science, Beijing University of Posts and Telecommunications, Beijing Key Laboratory of Network Systems and Network Culture, 100876, No.10 Xitucheng Road, Haidian District, Beijing, China 13811802242.

Zhang Bo, 1987.12, Shanxi, Chinese, Male, Master, HCI cognitive science, Beijing University of Posts and Telecommunications, Beijing Key Laboratory of Network Systems and Network Culture 100876, No.10 Xitucheng Road, Haidian District, Beijing, China 13811802242.

Instructions for Authors

Manuscript Submission

All paper submissions will be handled electronically in EDAS via the JMM Submission Page (URL: http://edas.info/newPaper.php?c=7325). After login EDAS, you will first register the paper. Afterwards, you will be able to add authors and submit the manuscript (file). If you do not have an EDAS account, you can obtain one. Along with the submission, Authors should select up to 3 topics from the EDICS (URL: http://www.academypublisher.com/jmm/jmmedics.html), and clearly state them during the registration of the submission.

JMM invites original, previously unpublished, research papers, review, survey and tutorial papers, application papers, plus case studies, short research notes and letters, on both applied and theoretical aspects. Submission implies that the manuscript has not been published previously, and is not currently submitted for publication elsewhere. Submission also implies that the corresponding author has consent of all authors. Upon acceptance for publication transfer of copyright will be made to Academy Publisher, article submission implies author agreement with this policy. Manuscripts should be written in English. Paper submissions are accepted only in PDF. Other formats will not be accepted. Papers should be formatted into A4-size (8.27" x 11.69") pages, with main text of 10-point Times New Roman, in single-spaced two-column format. Authors are advised to follow the format of the final version at this stage. All the papers, except survey, should ideally not exceed 12,000 words (14 pages) in length. Whenever applicable, submissions must include the following elements: title, authors, affiliations, contacts, abstract, index terms, introduction, main text, conclusions, appendixes, acknowledgement, references, and biographies.

Conference Version

Submissions previously published in conference proceedings are eligible for consideration provided that the author informs the Editors at the time of submission and that the submission has undergone substantial revision. In the new submission, authors are required to cite the previous publication and very clearly indicate how the new submission offers substantively novel or different contributions beyond those of the previously published work. The appropriate way to indicate that your paper has been revised substantially is for the new paper to have a new title. Author should supply a copy of the previous version to the Editor, and provide a brief description of the differences between the submitted manuscript and the previous version.

If the authors provide a previously published conference submission, Editors will check the submission to determine whether there has been sufficient new material added to warrant publication in the Journal. The Academy Publisher's guidelines are that the submission should contain a significant amount of new material, that is, material that has not been published elsewhere. New results are not required; however, the submission should contain expansions of key ideas, examples, elaborations, and so on, of the conference submission. The paper submitting to the journal should differ from the previously published material by at least 30 percent.

Review Process

Submissions are accepted for review with the understanding that the same work has been neither submitted to, nor published in, another publication. Concurrent submission to other publications will result in immediate rejection of the submission.

All manuscripts will be subject to a well established, fair, unbiased peer review and refereeing procedure, and are considered on the basis of their significance, novelty and usefulness to the Journals readership. The reviewing structure will always ensure the anonymity of the referees. The review output will be one of the following decisions: Accept, Accept with minor changes, Accept with major changes, or Reject.

The review process may take approximately three months to be completed. Should authors be requested by the editor to revise the text, the revised version should be submitted within three months for a major revision or one month for a minor revision. Authors who need more time are kindly requested to contact the Editor. The Editor reserves the right to reject a paper if it does not meet the aims and scope of the journal, it is not technically sound, it is not revised satisfactorily, or if it is inadequate in presentation.

Revised and Final Version Submission

Revised version should follow the same requirements as for the final version to format the paper, plus a short summary about the modifications authors have made and author's response to reviewer's comments.

Authors are requested to use the Academy Publisher Journal Style for preparing the final camera-ready version. A template in PDF and an MS word template can be downloaded from the web site. Authors are requested to strictly follow the guidelines specified in the templates. Only PDF format is acceptable. The PDF document should be sent as an open file, i.e. without any data protection. Authors should submit their paper electronically through email to the Journal's submission address. Please always refer to the paper ID in the submissions and any further enquiries.

Please do not use the Adobe Acrobat PDFWriter to generate the PDF file. Use the Adobe Acrobat Distiller instead, which is contained in the same package as the Acrobat PDFWriter. Make sure that you have used Type 1 or True Type Fonts (check with the Acrobat Reader or Acrobat Writer by clicking on File>Document Properties>Fonts to see the list of fonts and their type used in the PDF document).

Copyright

Submission of your paper to this journal implies that the paper is not under submission for publication elsewhere. Material which has been previously copyrighted, published, or accepted for publication will not be considered for publication in this journal. Submission of a manuscript is interpreted as a statement of certification that no part of the manuscript is copyrighted by any other publisher nor is under review by any other formal publication.

Submitted papers are assumed to contain no proprietary material unprotected by patent or patent application; responsibility for technical content and for protection of proprietary material rests solely with the author(s) and their organizations and is not the responsibility of the Academy Publisher or its editorial staff. The main author is responsible for ensuring that the article has been seen and approved by all the other authors. It is the responsibility of the author to obtain all necessary copyright release permissions for the use of any copyrighted materials in the manuscript prior to the submission. More information about permission request can be found at the web site.

Authors are asked to sign a warranty and copyright agreement upon acceptance of their manuscript, before the manuscript can be published. The Copyright Transfer Agreement can be downloaded from the web site.

Publication Charges and Re-print

The author's company or institution will be requested to pay a flat publication fee of EUR 360 for an accepted manuscript regardless of the length of the paper. The page charges are mandatory. Authors are entitled to a 30% discount on the journal, which is EUR 100 per copy. Reprints of the paper can be ordered with a price of EUR 100 per 20 copies. An allowance of 50% discount may be granted for individuals without a host institution and from less developed countries, upon application. Such application however will be handled case by case.

More information is available on the web site at http://www.academypublisher.com/jmm/authorguide.html.