Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®
Pattern Recognition Letters 24 (2003) 2913–2922
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Recognizing image "style" and activities in video using local features and naive Bayes

Daniel Keren *

Department of Computer Science, University of Haifa, Haifa 31905, Israel

Received 6 October 2002; received in revised form 13 May 2003

Abstract

The goal of this paper is to offer a framework for classification of images and video according to their "type", or "style"—a problem which is hard to define, but easy to illustrate; for example, identifying an artist by the style of his/her painting, or determining the activity in a video sequence. The paper offers a simple classification paradigm based on local properties of spatial or spatio-temporal blocks. The learning and classification are based on the naive Bayes classifier. A few experimental results are presented.

Keywords: Image style; Texture; Naive Bayes; Activity detection

1. Introduction

One of the visual tasks which humans perform well may be described as "recognition by style". For example, a person can quite successfully determine the identity of an artist given a hitherto not seen painting, if he/she is familiar with other paintings made by the artist, and if two artists with very similar styles are not present (still, in that case, it is possible to recognize the school of the painting—cubist, expressionist, etc.).

The recognition of "style" does not use gray level or color similarity, nor high-level features (such as faces, eyes, etc.), which excludes using many methods that are successful for other computer vision problems. Another interesting difficulty in the "style detection" problem is the construction of a training set, since, for example, every Dali painting is not "pure Dali", and it will have some areas in it which appear as if they were painted by, say, Van-Gogh. Hence, the training sets of the positive and negative examples respectively will contain some negative and positive elements. The non-linear nature of the suggested method overcomes this difficulty, and also allows it to handle the case in which different styles are mixed (averaged).

This paper offers a simple, fast, and very easy to implement algorithm. It chooses local features which are based on the DCT transform coefficients, and then classifies the image/video blocks using the *naive Bayes* classifier. It may be viewed as

* Tel.: +972-4-824-9730; fax: +972-4-824-9331.
E-mail address: dkeren@cs.haifa.ac.il (D. Keren).

0167-8655/$ - see front matter
doi:10.1016/S0167-8655(03)00152-1

a test as to how "very local" methods can identify style; in that regard, it is the opposite extreme of histogram-based methods such as the one described in (Zelnik-Manor and Irani, 2001). Such local methods are important for cases in which a few "styles" co-exist closely together in an image or video sequence.

2. Previous work

There is a large body of work related to the topic of this paper (which usually falls under the category of texture-based classification). The length of this paper allows to mention only some references to recent work. In (Bell and Freeman, 2001), a mixture model was fitted to the output of a filter bank to classify shading and reflectance variations. Tieu and Viola (2000) applies boosting to choose highly selective features for classification. A sophisticated non-parametric multi-scale texture analysis was presented in (de Bonet and Viola, 1998). An application of global coefficient statistics to noise removal was offered in (Simoncelli and Adelson, 1996). Images were classified by the rate of decay of their Fourier spectrum in (Voss, 1996). More recent work is presented in (Barnard and Forsyth, 2001; Belongie et al., 1998; Flickner et al., 1995; Greenspan et al., 1994; Thomson and Foster, 1997; Wang et al., 2001).

3. The naive Bayes classifier

The *naive Bayes* classifier is very popular in the data retrieval community, especially in text categorization applications (Dumais et al., 1998; Lewis, 1998). A short survey of the method's implementation follows:

1. A training set is given, which consists of a set of examples from the categories $\{C_1, C_2, \ldots, C_m\}$. Denote the number of C_i examples as n_i, and the total number as $\sum n_i = n$. The probability of the ith category is defined by $P(C_i) = n_i/n$. I shall refer to the examples as *texts*, although they do not necessarily have to be textual.
2. Define a set of possible *features*. In textual applications, these are usually words, classes of words which have a similar meaning, or "word stems". A feature may or may not appear in a text. For every feature f_i and category C_j, define $P(f_i/C_j)$ as the ratio of C_j's members which contain f_i, and $P(f_i)$ as the ratio of all members of all categories which contain f_i. The important notion of *mutual information* between a feature f_i and category C_j is defined as

$$\mathrm{MI}(f_i, C_j) = P(f_i/C_j) \log\left(\frac{P(f_i/C_j)}{P(f_i)}\right) \tag{1}$$

   The mutual information has an attractive intuitive meaning; for it to be large, the frequency of f_i in C_j has to large in absolute terms, and it also has to be large relative to f_i's frequency in all the categories (its average frequency).
3. For every category, choose a few features which have the largest mutual information with respect to it. The union of these sets over all categories is called the *feature set*.
4. Given a new text T, extract all the features which it contains—call them $\{f_{i_1}, f_{i_2}, \ldots, f_{i_k}\}$—and estimate for every category C_j the probability that T belongs to it, by

$$\begin{aligned} P(C_j/T) &= \frac{P(C_j)P(T/C_j)}{P(T)} \\ &\approx \frac{P(C_j)P(\{f_{i_1}, f_{i_2}, \ldots, f_{i_k}\}/C_j)}{P(\{f_{i_1}, f_{i_2}, \ldots, f_{i_k}\})} \\ &\approx \frac{P(C_j)\prod_{l=1}^{k} P(f_{i_l}/C_j)}{\prod_{l=1}^{k} P(f_{i_l})} \end{aligned} \tag{2}$$

   The first equality is just Bayes' law. The first approximation means that, when classifying T, I only consider the features it contains. The second approximation assumes that the presence of features is independent (this is where the "naive" in "naive Bayes" comes from); while this is not always true, the technique is still surprisingly effective.
5. Usually, the "non-events"—that is, the non-appearances of a feature in a text—are also considered, which leads to a straightforward extension of Eq. (2).

4. Applying the naive Bayes method to image classification

The first problem hindering the application of naive Bayes to image classification is: what are the analogues of "text" and "feature" in images? For the task of detecting images which contain some pre-defined structures, one may define a feature as a certain sub-image. For example, for detecting images with human faces, a useful feature would be the presence of an eye in the image. In (Ullman et al., 2001), such "informative features" were recovered, and various algorithms used to classify images based on the features' presence. Certain textures can also be recognized by the presence of templates, perhaps up to rotation or scale, etc. Such features, however, are unsuitable for the problem of style detection as presented here (unless I identify a painting by the artist's signature...). In general, one cannot hope to base "style classification" on the presence of a few features.

Instead, I offer to classify every image block, and then classify the entire image by a majority vote. The information extracted from this process contains more than the classification of the entire image; it maps the image to different regions, each dominated by a certain style. As will be demonstrated in Section 7, this often yields results which agree with human intuition. The local analysis of the image contains more information than that present in histogram-based approaches, which classify the entire image based on similarity between cumulative distributions of gradients, or wavelet coefficients, etc.

As opposed to the text categorization applications of naive Bayes, and also to Ullman et al. (2001), this paper suggests to use *features which have the same size as texts*. I treat each and every image block (the size in the experiments was 9×9, and sliding blocks were used) as a text, and the features are the 9×9 DCT basis functions. I say that a certain such feature (coefficient) appears in a block if its absolute value in the block's expansion is larger than a certain threshold; in Section 5 I explain how this threshold is determined (see also Fig. 2).

One may wonder how such small blocks can capture the style by which an artist draws. The best answer I can offer is that I am not sure; however, the algorithm does succeed to do this, to a reasonable degree. One possibility is that structures which seem "large" to us (for example, the "wavy" patterns in Van-Gogh's paintings) also exist on a smaller scale—my opinion is that this is indeed the case. Another possibility is that the algorithm captures small features which a human observer does not notice. This may be an interesting question to pursue. Let me also note that when the size of the blocks is increased to 18×18, the performance drops significantly.

5. Implementation

The suggested implementation of the classifier to the problem of "style" detection proceeds as follows (the explanation is presented to classification of paintings, but the algorithm is general):

1. Build an image database. Here, I have tested five artists—Rembrandt, Van-Gogh, Picasso, Magritte, and Dali. Ten paintings by each artist consisted the training set, and the test set consisted of 20–10 paintings for each artist. The training set was randomly chosen.

2. For each DCT basis element (9×9 in size), b_{ij}, and for every artist, the absolute values of the DCT coefficient corresponding to b_{ij} are computed for every 9×9 block in all the artist's paintings in the training set. These values are then binned into 100 discrete values. The blocks are first normalized to zero mean and unit variance, hence the absolute values of the coefficients are between 0 and 1. All these operations can be implemented using convolutions, hence can be done rather quickly. Then, it is straightforward to construct a table $T(p,i,j,a)$, which stores the probability that, for the artist p, the absolute value of the (i,j) DCT coefficient is greater or equal than a. Here p ranges over all artists, i and j between 1 and 9, and a ranges over $\{0, 0.01, 0.02, \ldots, 0.99, 1\}$.

3. Naive Bayes requires binary features (by that I mean features that either appear or do not appear, like words which may or may not appear in a document), so I have to convert the continuous presence of a basis element in a block

(that is, its coefficient in the block's expansion), to a binary one. This is done by thresholding the coefficient's absolute value. For every pair of artists and every coefficient, the threshold is chosen so as to maximize the mutual information (Eq. (1)); that is, I assume that there are only two categories—consisting of the paintings of these two artists—and find the optimal features for each of them. Note that this is a very fast process, once the probability table of stage 2 was built. The maximization is performed over each binned value $\{0, 0.01, 0.02, \ldots, 0.99, 1\}$, and over both artists.

Let me note that there are other ways in which the continuous information could be reduced to a binary one. I could, for example, use many "buckets" representing different ranges of the DCT values, thus increasing the number of features candidates. However, experience has shown that the distributions of the DCT coefficients for every painter very typically behave roughly like Gaussians. For this reason I chose the relatively simple and computationally efficient "binarization" by a single threshold, which attempts to separate the Gaussians of two different artists as best as possible. Admittedly, this may fail; suppose for example that there are many Dali blocks with a certain DCT coefficient in the interval $[0.1, 0.2]$ and many other Dali blocks for which that coefficient is in the interval $[0.6, 0.7]$, while the respective Van-Gogh coefficients are concentrated in [0.4, 0.5]. In that case, a single threshold will never do a good job in separating Dali and Van-Gogh. In my experience—and this is a purely empirical observation—this does not happen. However, for problems with a more complicated distribution of the coefficients, one can extend the algorithm to allow more features. A typical feature may be, for example, the presence of the coefficient not only in an interval $[0, a]$ or $[b, 1]$, but in a union of two disjoint intervals. This will, naturally, considerably increase the number of candidate features and the computational cost of finding the best features for discrimination.

4. For each artist in each pair, the ten features with the highest mutual information are chosen. Note that each feature consists of a basis element and a threshold for its coefficient in a block's expansion.

5. Given a new image and a pair of artists, the probability of each image block with respect to each artist is computed from Eq. (2). Better results were obtained by considering only blocks with a (pre-normalization) variance higher than a certain threshold—20 was a good value, but the results do not change much if 10 or 30 is used. Another heuristic which yielded better results was to classify only blocks for which the winning artist's probability was at least twice the other artist's probability.

I note here that these heuristics do not have a substantial influence on the algorithm's performance. Such questions—the number of features to use, etc.—are inherent to naive Bayes and other classification methods, and these are difficult and yet unanswered questions. The main point of this short paper is the novel application of naive Bayes, and, hopefully, future work will address these heuristics more rigorously.

6. Every pixel in the test image is assigned a label, according to the classification of the 9×9 block surrounding it. Each individual pixel is thus classified (not all the window pixels). The boundary pixels are discarded (since the windows are 9×9 and the images much larger, this is not much of a problem). Pixels whose corresponding window's variance is too small, or for which the ratio between the large and small probabilities does not exceed 2, are labeled as unclassified.

Since every DCT coefficient is equal to the inner product of the respective basis element with the current window, the computation can be implemented by convolution with the basis element, and is quite efficient. It takes far less than a second to classify all the pixels of a large image.

7. The overall classification is determined by a majority vote. However, as noted before, the mapping of individual pixels to different artists contains more information than the overall classification.

5.1. Why DCT?

Recall that naive Bayes assumes that the features are independent. Since the DCT basis elements are orthogonal, if $(i_1, j_1) \neq (i_2, j_2)$ and b_{i_1,j_1}, b_{i_2,j_2} are the respective basis elements, then under the "natural" probability distribution over

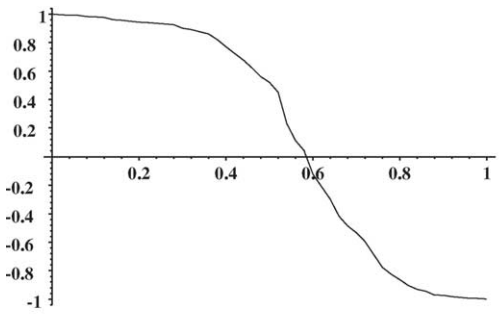
images (Keren and Werman, 1993), the random variables $I \to (I, b_{i_1,j_1})$ and $I \to (I, b_{i_2,j_2})$ (where I varies over images), are independent random variables over the space of *all* images.

One may rightly question the validity of the independence of the DCT coefficients, which may be false in a "small world", corresponding to a single artist for example. Therefore, a decorrelated basis was created by diagonalizing the covariance matrix for each artist, and classification according to this basis was attempted. Somewhat to the author's surprise, this did not result in any substantial improvement. A possible explanation as to why naive Bayes can overcome feature dependencies is offered in (Lewis, 1998).

6. Classifying a mixture of styles

Suppose we are given a mixture (in the form of an average) of images of different styles. Let me note here that simple averaging may not necessarily be a good model for an artist who paints in a "mixed style"; it is perhaps more suitable when transparencies exist in the scene (e.g. looking into a room via a window in which greenery is reflected). If the illumination model is multiplicative, we can remedy this by taking logarithms (as in homomorphic filtering). But, regardless of the possible application, I demonstrate the results for paintings, with the hope of convincing that it is applicable to other problems as well.

It may be expected that, at least as far as human intuition is concerned, the classification of the mixture will not be linear; that is, the classification of a weighted average of styles will be biased towards the dominant one. To test the behavior of the classifier on a mixture of styles, weighted averages of two paintings—a self-portrait by Van-Gogh (P_1) and "Metamorphosis of Narcissus" by Dali (P_2)—were created (after the images were normalized to the same size). The mixtures varied over convex combinations, $\lambda P_1 + (1-\lambda)P_2$, for $\lambda = \frac{k}{100}$, $0 \leqslant k \leqslant 100$. For each λ, the mixture image was classified. A numerical measure assigned to each λ was defined as the difference between the number of "Dali blocks" and "Van-Gogh blocks". The two "pure" images ($\lambda = 0, 1$) were normalized to 1 and -1 respectively. Fig. 1 depicts the relation between λ and this measure.



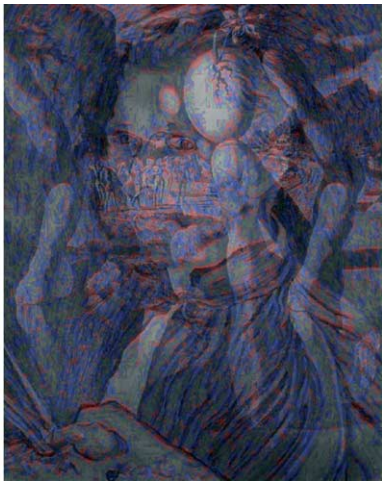


Fig. 1. Left: strength of classification as a measure of λ (which is represented by the horizontal axis), of the mixture $\lambda P_1 + (1-\lambda)P_2$, where P_1 stands for Van-Gogh and P_2 for Dali. Classification strength is determined by the ratio of P_1 vs. P_2 blocks. The strength is linearly normalized such that the cases $\lambda = 0$ (pure Dali) and $\lambda = 1$ (pure Van-Gogh) are assigned strengths of 1, -1 respectively. Results are consistent with human intuition which is biased towards the dominant style. This non-linear behavior, which results in the classification being robust under "contamination" by another style, is a result of the non-linear thresholding of the DCT coefficients. Right: one of the images used in the mixture classification. "Dali pixels" are reddish, "Van-Gogh" bluish. Here and elsewhere, unclassified pixels are in gray level.

7. Results

For the five artists tested, a "tournament scheme" classifier was implemented (Pontil and Verri, 1998). The rate of success was 86%. Some examples are presented in Figs. 2–4.

7.1. Classifying images as "old" or "new"

Some experiments were made in discriminating old photographs (19th century) and photographs captured by a digital camera; the results are displayed in Fig. 5.

8. Activity detection in video

The "style classification" method was extended to detect activities in video sequences. This topic is drawing a lot of interest in recent years (Black et al., 1997; Bobick and Davis, 2001; Fleet et al., 2000). The literature is too numerous to cover in this short paper; for a good survey, one may consult Moeslund and Granum (2001). Temporal texture and gradient distribution approaches,

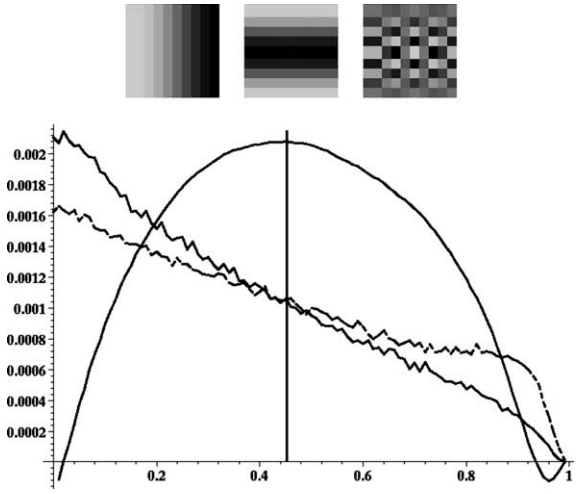


Fig. 2. Top: three DCT basis elements with highest mutual information for discriminating Dali from Van-Gogh. The mutual informations (left to right) were 0.042, 0.037, 0.036. Bottom: the distribution and mutual information for various thresholds of the DCT coefficient corresponding to the most discriminating feature (top left). The mutual information for thresholds between 0 and 1 is the parabola-like curve (empty circles); it is scaled by 0.05 for visualization purposes. The percentage of Dali blocks with the corresponding DCT coefficient (after binning) is the dotted line, and the solid line depicts the same for Van-Gogh blocks. The optimal threshold (for which the highest mutual information is obtained) is 0.45 (solid vertical line). Note that it is achieved roughly at the point in which the Dali and Van-Gogh curves intersect.

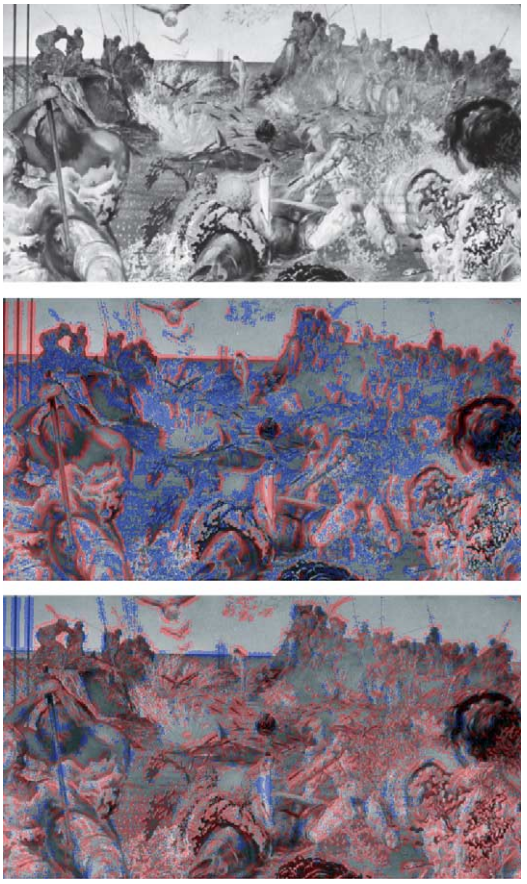


Fig. 3. Top: excerpt from Dali's "Tuna Fishing". Here and elsewhere, the classification was performed after the painting was transformed to gray levels. Middle: same excerpt after Dali/Van-Gogh classification; "Dali pixels" are reddish, "Van-Gogh pixels" bluish. In accordance with human intuition, "wavy" areas are dominantly Van-Gogh (for example, the part of the arm at the bottom right corner). Bottom: same excerpt after Dali/Magritte classification; "Dali pixels" are reddish, "Magritte pixels" bluish. Note, for example, that the vertical sharp structures in the top left are classified as Dali when compared against Van-Gogh, but as Magritte when compared against Magritte. This can be intuitively explained by the observation that "Dali paints with more straight lines than Van-Gogh but with less straight lines than Magritte".

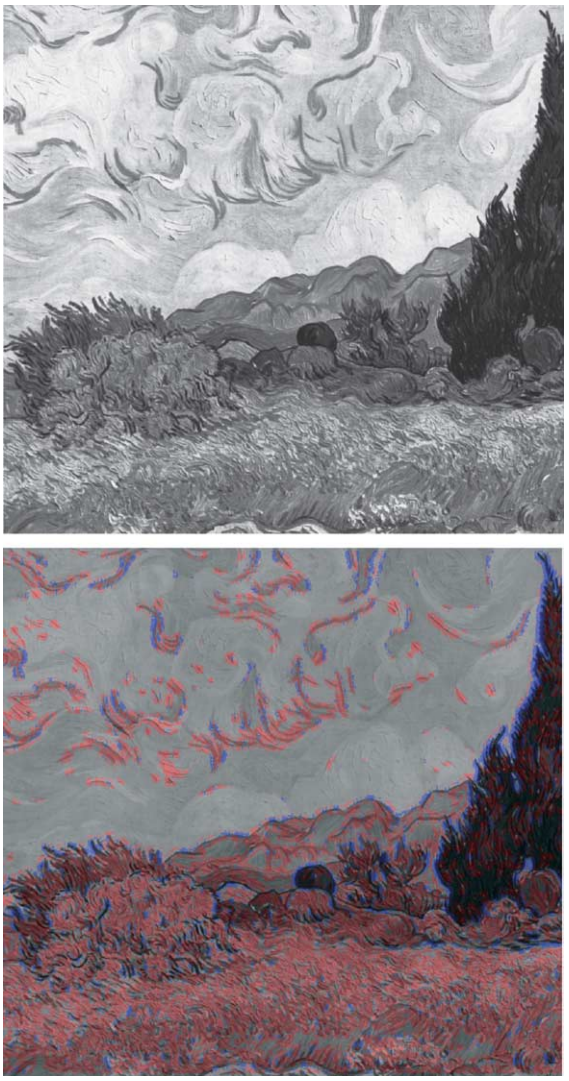


Fig. 4. Top: the painting "Wheat Field with Cypress" by Van-Gogh. Bottom: after classification vs. Dali. Van-Gogh pixels are reddish, Dali pixels bluish. The typical Van-Gogh wavy structures are correctly classified in general.

which can be viewed as an extension of texture classification in the spatial domain, are discussed in (Polana and Nelson, 1994, 1997; Zelnik-Manor and Irani, 2001).

In order to recover features which characterize a certain type of activity, the algorithm in Section 5 was extended to the spatio-temporal domain. First, three-dimensional "image stacks" are built from the movie segments in the training set. This is accomplished by constructing a three-dimensional array A, with $A[i,j,k] =$ the (i,j) pixel in the kth frame. In order to save memory, this can be done sequentially, with the probability table (stage 2 in Section 5) built by adding the data for subsequences. Instead of two-dimensional blocks, three-dimensional blocks are used, with the 3D DCT transform coefficients as features. A block B corresponding to the pixel with spatial coordinates (x,y) at time t consists of the pixels $B[i,j,k] = A[x+i, y+j, t+k]$, where i,j,k range in some small intervals centered at 0 (I used $5\times5\times5$ blocks). Blocks with a small time derivative (i.e. in which not much activity occurs) are not considered. This limits the algorithm to a stationary camera. If the camera is moving, stabilization and motion compensation can be used. The experiments were so far limited to distinguishing between two types of activity: hand waving and walking. Four different individuals were filmed while performing these activities, in three different locations: opposite a white wall, in an office, and in a corridor. Part of the captured video sequences were used for training, and then the classification algorithm was applied to the remaining sequences. A resolution of 64×64 was used, which is low enough to allow real-time classification (Figs. 6 and 7).

9. Conclusion and future research

A simple and very fast algorithm for image and activity "style" classification using the naive Bayes classifier was presented, and applied to the problems of artist identification, classifying photographs as digital or old, and activity detection in video (walking vs. hand waving). Further research will consist of incorporating a multi-level scheme and developing methods to determine the correct block size(s), as well as testing other representations than the DCT, such as wavelets and overcomplete bases. A Markov random field paradigm may be applied in order to create a more consistent segmentation of the image (i.e. not to allow an isolated "Dali pixel" in a "Van-Gogh area" of the image, as well as to preserve continuity of action

Fig. 5. Top: four of the images used as a training set for ''old pictures''. Bottom: result of classification for a photograph of Robert E. Lee. Pixels classified as ''old photograph'' pixels are reddish, ''digital camera'' are bluish.

Fig. 6. Eight frames from a low-resolution video sequence depicting a person walking across a corridor. If the spatio-temporal $5\times5\times5$ neighborhood of a pixel was classified as ''walking'' it was colored purple, and if it was classified as ''hand waving'' it was colored yellow. Most of the misclassification occurs in areas in which the diagonal motion of the legs resembles the upwards or downwards motion of the hands in the ''hand waving'' sequence (see Fig. 7). Some of the person's reflection is also classified as ''walking''. Altogether, 83% of the classified pixels were labeled as ''walking''.

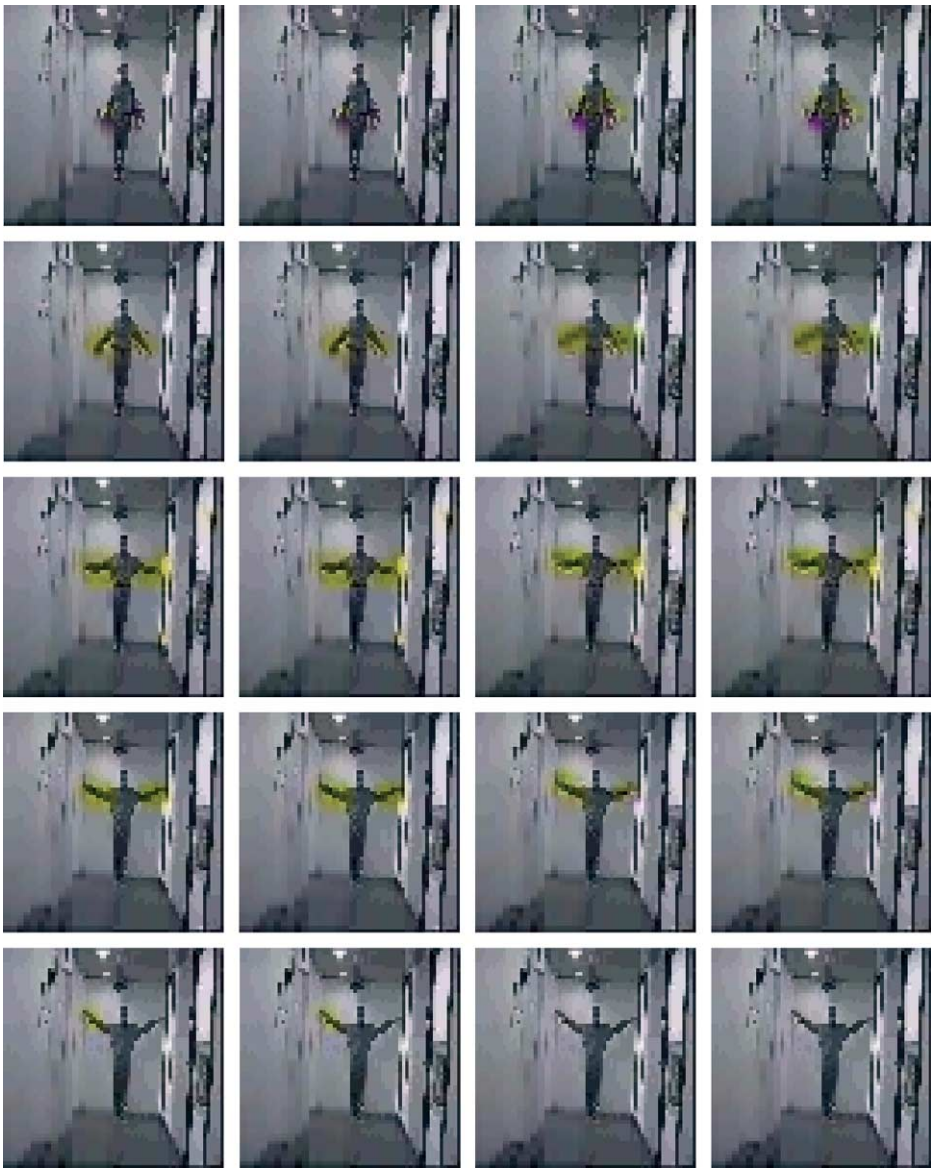


Fig. 7. Twenty frames from a low-resolution video sequence depicting a person waving his hands. If the spatio-temporal $5\times5\times5$ neighborhood of a pixel was classified as "walking" it was colored purple, and if it was classified as "hand waving" it was colored yellow. Most of the misclassification occurred within the first four frames, in which the diagonal motion of the hands at a low inclination locally resembles the leg motion of a walking person. Barely any motion was detected in the last four frames, because they depict the stage in which the upwards motion of the hands ends and there's a slight pause before the downwards motion commences. Altogether, 98% of the classified pixels were labeled as "hand waving".

over time), as well as additional high-level processing and grouping.

The advantages of the presented algorithm are in its simplicity and speed, and its ability to handle a large number of features. The results are reasonable and consistent with human intuition. Reasonable results were also obtained for video sequences although the resolution was rather low.

The suggested method is very local in nature and thus can handle a few different styles or activities which co-exist in an image or a video sequence. It can also overcome considerable "contamination" of one style by another, as demonstrated in Section 6.

Acknowledgements

This research was supported by The Israel Science Foundation (grant no. 591/00-10.5).

References

Barnard, K., Forsyth, D.A., 2001. Learning the semantics of words and pictures. In: International Conference on Computer Vision, II, pp. 408–415.

Bell, M., Freeman, W.T., 2001. Learning local evidence for shading and reflectance. In: International Conference on Computer Vision, I, pp. 670–677.

Belongie, S., Carson, C., Greenspan, H., Malik, J., 1998. Color- and texture-based image segmentation using the expectation-maximization algorithm and its application to content-based image retrieval. In: International Conference on Computer Vision, pp. 675–682.

Black, M.J., Yacoob, Y., Jepson, A.D., Fleet, D.J., 1997. Learning parameterized models of image motion. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 561–567.

Bobick, A.F., Davis, J.W., 2001. The recognition of human movement using temporal templates. IEEE Trans. Pattern Anal. Machine Intell. 23 (3), 257–267.

de Bonet, J.S., Viola, P.A., 1998. Texture recognition using a non-parametric multi-scale statistical model. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 641–647.

Dumais, S.T., Platt, J., Hecherman, D., Sahami, M., 1998. Inductive learning algorithms and representations for text categorization. In: Proceedings of the 7th International Conference on Information and Knowledge Management, pp. 148–155.

Fleet, D.J., Black, M.J., Yacoob, Y., Jepson, A.D., 2000. Design and use of linear models for image motion analysis. Internat. J. Comput. Vision 36 (3), 169–191.

Flickner, M.D., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P., 1995. Query by image and video content: The qbic system. Computer 28 (9), 23–32.

Greenspan, H., Goodman, R., Chellappa, R., Anderson, C.H., 1994. Learning texture-discrimination rules in a multiresolution system. IEEE Trans. Pattern Anal. Machine Intell. 16 (9), 894–901.

Keren, D., Werman, M., 1993. Probabilistic analysis of regularization. IEEE Trans. Pattern Anal. Machine Intell. 15 (October), 982–995.

Lewis, D., 1998. Naive Bayes at forty: The independence assumption in information retrieval. In: Proceedings of the 10th European Conference on Machine Learning, pp. 4–15.

Moeslund, T.B., Granum, E., 2001. A survey of computer vision-based human motion capture. Computer Vision and Image Understanding 81 (3), 231–268.

Polana, R., Nelson, R.C., 1994. Detecting activities. In: International Conference on Pattern Recognition, A, pp. 815–818.

Polana, R., Nelson, R.C., 1997. Temporal texture and activity recognition. In: Shah, M., Jain, R. (Eds.), Motion-Based Recognition, Kluwer-Academic, 1997 (Chapter 5).

Pontil, M., Verri, A., 1998. Support vector machines for 3D object recognition. IEEE Trans. Pattern Anal Machine Intell. 20 (6), 637–646.

Simoncelli, E.P., Adelson, E.H., 1996. Noise removal via Bayesian wavelet coring. In: IEEE International Conference on Image Processing, I, pp. 379–382.

Thomson, M.G.A., Foster, D.H., 1997. Role of second-order and third-order statistics in the discriminability of natural images. J. Opt. Soc. Am., A 14 (9), 2081–2090.

Tieu, K., Viola, P., 2000. Boosting image retrieval. In: IEEE Conference on Computer Vision and Pattern Recognition, I, pp. 228–235.

Ullman, S., Sali, E., Vidal-Naquet, M., 2001. Fragment-based approach to object representation and classification. In: Visual Form 2001, 4th International Workshop on Visual Form, IWVF-4, Capri, May 2001. In: Lecture Notes in Computer Science, vol. 2059. Springer, pp. 85–102.

Voss, R.F., 1996. Local connected fractal dimension analysis of early Chinese landscape paintings and X-ray mammograms. In: Fractal Image Encoding and Analysis. In: A NATO ASI Series Book. pp. 279–297.

Wang, J.Z., Li, J., Wiederhold, G., 2001. Simplicity: Semantics-sensitive integrated matching for picture libraries. IEEE Trans. Pattern Anal Machine Intell. 23 (9), 947–963.

Zelnik-Manor, L., Irani, M., 2001. Event-based analysis of video. In: IEEE Conference on Computer Vision and Pattern Recognition, II, pp. 123–130.