Image Emotional classification: Static vs. Dynamic

Wang Wei-ning, Yu Ying-lin and Zhang Jian-chao
Department of Electronic and Information Engineering
South China University of Technology
Guangzhou, P.R.China
wnwang@scut.edu.cn, ecylyu@scut.edu.cn

Abstract *- Grouping images into emotional categories is an important and challenging problem in content-based image retrieval. In this paper, we propose an approach to classify art paintings into emotional categories (dynamic vs. static). The key points are feature selection and classification algorithm. According to the strong relationship between notable lines of image and human sensations, a novel feature vector WLDLV (Weighted Line Direction-Length Vector) is proposed, which includes both orientation and length information of lines in an image. Then classification is performed by SVM (Support Vector Machine) and images can be classified into dynamic vs. static. Experimental results demonstrate the effectiveness and superiority of our approach.*

Keywords: Image classification, emotional classification, Support Vector Machine (SVM), Weighted Line Direction-Length Vector (WLDLV), dynamic, static.

1 Introduction

Grouping images into semantic categories is an important and challenging problem in content-based image retrieval. Some researchers realized specific image classifications such as indoor vs. outdoor [1], and city vs. landscape [2][3]. On the other hand, psychological studies have suggested that different images induce different emotions. However, most of current image processing and applications overlook the effects of emotions. Recently, Colombo et al. [4] proposed an innovative method to get high-level representation of images and videos. In their work, low-level information is aggregated according to a set of rules, and then transformed into emotional level phrases. Japanese researchers have attempted to build image retrieval systems by impression words [5][6][7] and made positive results. Although semantic description and classification in emotional way have become remarkable in recent years, the study in this field is still at the very beginning.

In this paper, we propose an approach to classify the art paintings into particular emotional classes (static vs. dynamic) through low-level image features. According to the strong relationship between lines of image and human sensations, a novel feature vector WLDLV (Weighted Line Direction-Length Vector) is proposed, which is computed based on edge detection through wavelet transform. WLDLV includes both orientation and length information of lines in an image. With WLDLV, images can be classified into dynamic vs. static classes using an SVM (Support Vector Machine).

The rest of this paper is organized as follows. The feature extraction scheme is described in detail in Section 2. Section 3 provides a brief review of Support Vector Machine (SVM). Experimental results comparisons and analysis are presented in Section 4 followed by the conclusion and future work in Section 5.

2 Feature Extraction

2.1 Relationship between line directions and dynamic sensation of images

It is a challenge to select good features for image classification. Studies have suggested that lines induce emotional effects [8][9]. For example, horizontal lines always associate with static horizon and communicate calmness and relaxation; vertical lines are clear and direct, and communicate dignity and eternality; slant lines are unstable and communicate notable dynamism. Lines with different directions can express different emotions, especially the longer, thicker and straighter one.

(a) (b) (c)

(d) (e) (f)

Figure 1. Example of static image and its edges and EDH. (a)&(d) are Original images; (b)&(e) show their edges; (c)&(f) show their edge direction histograms

(a) (b) (c)

Figure 2. Example of dynamic image and its edges and EDH. (a) is original image; (b) shows its edges; (c) show its edge direction histogram

Psychological survey is conducted to classify 200 images into two emotional classes (static and dynamic). Among the 200 images, 110 images are selected and divided into two classes, static class with 70 images and dynamic class with 40 images. Then edges are used to extract lines of an image because they contain valuable information of the object's boundary and shape, and the EDH (Edge Direction Histogram)[2][3] is taken as the line direction distribution of an image. Static images (see Figures 1 (a) (d)) are characterized by vertical and horizontal lines, such as skylines, horizon and contours of tree trunks or other static objects. In their EDH, the number of edge points near horizontal or vertical directions is bigger than that of other directions. However, dynamic images (see Figure 2 (g)) are characterized by slant lines, such as contours of moving people or flying birds. A large proportion of edge points are in slant directions. It shows that line direction has a basic contribution to the static or dynamic sensation.

Table 1. Mean of Hor_vert and Slant for static and dynamic images

	Static images	Dynamic images	t-test
Mean of Hor_Vert	0.4412	0.3750	5.6796
SD of Hor_Vert	0.0641	0.0482	
Mean of Slant	0.5636	0.6305	5.7628
SD of Slant	0.0640	0.0476	

Statistical analysis shows the relevancy between line direction distribution and image sensation of static or dynamic in Table 1. For each class (static or dynamic), mean of Hor_Vert, which is the ratio of edge points near horizontal and vertical directions with respect to the overall number of edge points in the image, and mean of Slant, which is the ratio of edge points near 45° and 135° directions, are defined as the statistical variables to estimate the relevancy. The statistical results of 70 static images and 40 dynamic images are displayed in Table 1. It shows that the mean of Hor_Vert of the static images is larger than that of dynamic images, while the mean of Slant is just the reverse. In order to determine whether this difference is statistically significant or not, we conduct the t-test, which is a common statistic method to decide whether the difference of mean values is significant. Hypothesize that the mean values of Hor_vert of static images and dynamic images are not different, and the t-test value is calculated. The result of the t-test is 5.6796, thus the hypothesis is rejected at a 99.995% according to the t-Distribution. The same result is gotten to the Slant. This means that line direction distribution is a significant character to discriminate the two classes.

In order to fully describe line information, new feature WLDLV is introduced for static vs. dynamic classification.

2.2 WLDLV (Weighted Line Direction-length Vector) generation

According to the strong relationship between notable lines of images and human sensations, a new feature vector WLDLV (weighted line direction-length vector) is proposed, which includes both orientation and length information of lines in an image, and is computed by the following five steps:

Edge detection: Multi-scale edge detection can be achieved through a wavelet transform [10]. The angles of edge points can be determined at each scale. In our experiment, quadratic spline wavelet of compact support which is continuously differentiable [10] is selected to perform edge detection. The edge orientation of the image is gotten at the same time.

Edge orientation quantization: The corresponding edge directions are quantized into 4 segments: $(-15° \le \theta < 15°)$, $(15° \le \theta < 75°)$, $(75° \le \theta < 105°)$, $(105° \le \theta < 165°)$. An orientation vector H with 4 elements is computed where H(1) represents the number of horizontal edge pixels, H(3) is the number of vertical edge pixels and H(2), H(4) represent the numbers of slant edge pixels.

Histogram rotate: Figure 1(a) is an image which has more horizontal lines and less vertical lines, while Figure 1(d) is the reverse. Although both horizontal and vertical lines have the same effect on static sensation, the vector distance between Figure 1(a) and Figure 1(d) is bigger than we expected. Therefore, we shift the vector of Figure 2 for 90° to reduce the intra-class distance of static images. We take this step on the vectors whose H(3) is bigger than H(1).

Histogram split: The LDLV (Line Direction-Length Vector) is constructed in this stage. We compute the size of every connected component of edges in every direction segment. If the size of an edge component is less than K1 (K1=7), the pixels of the edge component are defined as short-line pixels; if the size is larger than K2

(K2=20), the pixels are defined as long-line pixels; and middle-line pixels otherwise. In this way, an orientation vector H splits into a two-dimensional array (matrix), consisting of three rows and four columns. The (i, j) elements of this matrix ($1 \le i \le 3$, $1 \le j \le 4$) indicates the number of edge points in i length lines (i=1 represents short-line, i=2 represents middle-line and i=3 represents long-line) with the orientation j.

WLDLV: Edge length is related with emotions. The longer the line is, the stronger the emotion will be deduced. Therefore, different weights are assigned to the lines according to their length. The longer the line, the bigger the weight is. In our experiments, w1, w2, and w3 are used for short-line, middle-line, long-line respectively (w1=0.1, w2=0.4, w3=0.5). Finally we normalize the WLDV and use it as input vector for classifier.

In this way, a feature vector WLDLV with 12 elements is generated for each image and will be used as the input vector for classifier in Section 3. Figure 3 shows the examples of WLDLV of Figure 1 (a) (d) and Figure 2 (a) images. Generally, static images have bigger ratios in orientation j=1 and 3, while dynamic images have bigger ratios in orientation j=2 and 4.

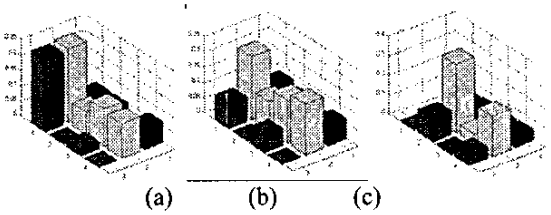


Figure 3. Examples of WLDLVs

(a), (b), (c) are corresponding to the original images of Figure 1. (a), (d), Figure 2.(e)

Using the Euclidean distance as a distance measure for the feature WLDLV, we generated the intra-class distance and the inter-class distance for the 110 images. Table 2 shows that the mean of intra-class distance is lower than that of inter-class distance. T-test is performed and the result shows that the difference is statistically significant. This means that the feature WLDLV has sufficient discrimination ability for static vs. dynamic classification.

Table 2. Intra-class and Inter-class distance for static and dynamic images

	Intra-class	Inter-class
Mean	0.2308	0.2556
SD	0.0860	0.0843
t-test	11.2882	

3 Classification by SVM

The support vector machine (SVM) approach is considered as a good classifier because it can achieve high generalization performance without the need of priori knowledge, even when the dimension of the input space is very high. An SVM is an approximate implementation of the structural risk minimization (SRM) method. It creates a classifier with the minimized Vapik-Chervonenkis (VC) dimension [11][12].

For a simple binary classification problem, given a set of training vectors belonging to two separate classes: $(x_1, y_1)\ (x_2, y_2) \ldots (x_N, y_N)$, $x_i \in R^n$, $y_i \in \{-1, +1\}$. If the two classes are linearly separable, the hyper plane that does the separation is

$$\boldsymbol{w} \cdot \boldsymbol{x} + b = 0 \tag{1}$$

The goal of an SVM is to find the parameters w_0 and b_0 for an optimal hyper plane to maximize the distance between the hyper-plane and the closest data point. This linear classifier is termed the Optimal Separating Hyperplane (OSH). Since the distance to the closest point is $1/\| \boldsymbol{w} \|$, finding the OSH amounts to minimize $\| \boldsymbol{w} \|$ and the objective function is:

$$\text{Minimize: } \phi(\boldsymbol{w}) = \frac{1}{2} \| \boldsymbol{w} \|^2 = \frac{1}{2}(\boldsymbol{w} \cdot \boldsymbol{w}),$$

$$\text{Subject to: } y_i (\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \ge 1 \quad i=1,\ldots,N \tag{2}$$

The constrained optimization problem can be achieved by using quadratic programming. The classification function can be written as

$$f(\boldsymbol{x}) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x}_i \cdot \boldsymbol{x} + b\right) \tag{3}$$

where b is the bias and the coefficients α_i are the Lagrange multipliers associated with constraint:

$$\alpha_i \ge 0\ ,\ \sum_{i=1}^{N} \alpha_i y_i = 0 \tag{4}$$

In the presence of noise, SVM introduced slack variables $\xi_i \ge 0$, and set an upper bound on the size of the α_i. Therefore, the Lagrange multipliers are modified to $0 \le \alpha_i \le C$, i=1,...N, where C is a penalty factor to penalize training errors. The choice of C is not strict in practice, and we set C=100 in our experiments.

If the two classes are nonlinearly separable, the input vectors should be nonlinearly mapped to a high

dimensional feature space by an inner-product kernel $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$, when the kernel satisfies Mercer's condition. Kernels commonly used include polynomials kernel, Gaussian RBF kernel and sigmoid kernel etc. Using different kernels, SVMs implement a variety of learning machines. We choose the Gaussian radial basis function in our experiments, which has the form:

$$k(x, x_i) = \exp(-\frac{\| x - x_i \|^2}{2\sigma^2}) \quad (5)$$

where σ is the width of the Gaussian function.

4 Experiments and Results

In order to achieve the best experimental results and algorithm generalization ability, large sample images are preferred. However, people's emotions are changed by time, especially after a long time test. To get the exact result, the estimate process should be finished within acceptant time duration, thus the number of sample images is limited. As a trade-off, 200 sample images are selected, and observers need one or two hours to finish the test. Due to sample capacity, SVM is selected because it can achieve excellent classification ability and generalization ability on limited samples.

Static and dynamic sensations are subjective emotions, and observers may get different results from the same image, for example, somebody thinks an image is static while the others think it is dynamic. In this paper, our approach is to get an algorithm which can classify images as human being. Since people can't get agreement with the class of the amphibolous images, the images will never be classified correctly by an algorithm. To get a meaning classification, the amphibolous images are excluded.

At first, 200 color art images were collected from painting gallery, and six undergraduates were invited to estimate the sensation of static or dynamic induced by the images. After excluding the amphibolous images, 110 color art images are gotten, which contain 70 static images and 40 dynamic images.. Each image is scaled to 192*288 or 288*192. WLDLV of each image is computed afterwards. In all cases, classification is performed using an SVM classifier with the Gaussian kernel.

The 110 images are separated into two set: training set with 80 images and testing set with 30 images. 40, 60 and 80 images of training set are used to train the SVM respectively. At the same time, the effectiveness of WLDLV is compared with another vector Edge Direction Histogram(EDH) which has shown effectiveness in image classification such as city vs. landscape classification [2][3]. The performance of EDH and WLDLV are shown in Table 3.

Table 3. Performance of EDH and WLDLV with different size of training set

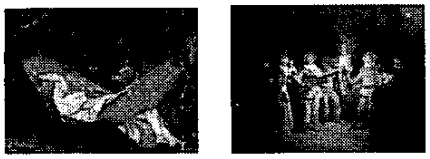
The size of training set		80 (100%)	60 (75%)	40 (50%)
WLDLV	AR(Accuracy Rate)	**93%**	91%	88%
	AR of static class	99%	98%	98%
	AR of dynamic class	80%	78%	68%
EDH	AR(Accuracy Rate)	**73%**	73%	71%
	AR of static class	80%	82%	83%
	AR of dynamic class	60%	55%	46%

Based on the experimental result, the following key points are demonstrated: (1) The accuracy rate increases as the number of the training images increases; (2) Although the dimension of the EDH (36) is higher than that of WLDLV (12), WLDLV got a much higher accuracy rate. WLDLV does a good job in discriminating images of two categories. (3) In all cases, the accuracy rate of dynamic class is lower than that of static class. The reason is that still-life and landscapes appear more frequently than dynamic content in paintings, hence the number of static images is bigger than that of dynamic images in our database.

Figure 4 shows some examples of the experimental results. Analyzing the experimental results, we find that 2 kinds of images are likely to be misclassified: (1) Images with static objects whose edges are slant lines. (2) Images with dynamic objects whose edges are close to horizontal or vertical lines. The misclassification problems are caused by the nature limit of WLDLV and should be solved by adding proper features.

(a) The right classified static images

(b) The right classified dynamic images



(c) static image misclassified to dynamic (d) dynamic images misclassified to static

Figure 4. Examples of the experimental images

5 Conclusions

In this paper, we propose a new feature WLDLV and use SVM for classification of image emotional semantics (static vs. dynamic). Experimental results show that WLDLV has a sufficient discrimination ability for this classification. Because of the subjectivity and complexity of the emotions themselves, WLDLV still has its limitation. To control the subjectivity and complexity, the experiment should be refined and contains more sample images. Adding other features into the input vector of emotional classification process may improve the performance of classification.

In future work, we will consider these problems more detailedly, such as adding other features (including color, texture and so on), investigating emotional effects of these features, finding feasible ways of feature extraction and reasonable description of the emotional semantics and using efficient emotional classification approaches.

Acknowledgements

The authors would like to acknowledge the support from National Natural Science Foundation of China (NO. 60372068)

References

[1] M.Szummer, and R. W. Picard, "Indoor-outdoor image classification" In IEEE International Workshop on Content-based Access of Image and Video Databases, India, pp.42-102, Jan 1998.

[2] A. Vailaya, A. K. Jain, and H. J. Zhang, "On Image Classification: City vs. Landscape," in IEEE Workshop on Content-Based Access of Image and Video Libraries, (Santa Barbara, CA), June 21, 1998, pp. 3–8.

[3] A. Vailaya, M. Figueiredo, A. K. Jain, and H. J. Zhang "A Bayesian Framework for Semantic Classification of Outdoor Vacation Images" IEEE Transactions on Image Processing , vol.10, No.1, pp.117-130, 2001.

[4] C. Colombo, A. Del Bimbo, and P. Pala, Semantics in Visual Information Retrieval[J], IEEE Multimedia, vol. 6, No.3, pp. 38-53, 1999.

[5] Hayashi. T., Hagiwara, M., "Image query by Impression words—The IQI System" IEEE Transactions on Consumer Electronics, vol.44, No.2, pp.347 -352, 1998.

[6] Yoshida, K., Kato, T., Yanaru, T., "Image Retrieval System Using Impression Words Systems", IEEE International Conference on Man, and Cybernetics, vol.3, No.11-14, pp. 2780 -2784, 1998.

[7] N. Bianchi-Berthouze, C. Lisetti, "Modeling multimodal expression of users' affective subjective experience", International Journal on User Modeling and User-Adapted Interaction: Special Issue on User Modeling and Adaptation in Affective Computing, vol.12, No.1, pp.49-84, 2002.

[8] Jonhannes Itten, The Art of Color (Chinese version), Shanghai People's Art Press, Shanghai, 1992.

[9] Johannes Itten, Design and Form—The Basic Course at the Bauhous (Chinese version), Shanghai People's Art Press, Shanghai, 1992

[10] S.Mallat and S. Zhong, "Characterization of Signals from Multiscale Edges", IEEE Trains. Pattern Analysis Machine Intelligence, vol. 11, no.7, pp. 710-732, 1992.

[11] V.Vapnik. The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995

[12] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Date Mining and Knowledge Discovery, 2(2), pp.1-47, 1998.