Vision-based Hand Gesture Recognition Using PCA+Gabor Filters and SVM

Deng-Yuan Huang[1], Wu-Chih Hu[2], Sung-Hsiang Chang[3]
[1,3]Department of Electrical Engineering, Da-Yeh University
[1]kevin@mail.dyu.edu.tw, [3]r9603025@mail.dyu.edu.tw
[2]Department of Computer Science and Information Engineering, National Penghu University
wchu@npu.edu.tw

Abstract—**In this paper we present a novel method for hand gesture recognition based on Gabor filters and support vector machine (SVM). Gabor filters are first convolved with images to acquire desirable hand gesture features. The principal components analysis (PCA) method is then used to reduce the dimensionality of the feature space. With the reduced Gabor features, SVM is trained and exploited to perform the hand gesture recognition tasks. To confirm the robustness of the proposed method, a dataset with large posed-angle (>45 deg.) of hand gestures is created. The experiment result shows that the recognition rate of 95.2% can be achieved when SVM is used. A real-time video system for hand gesture recognition is also presented with a processing rate of 0.2 s for every frame. This result proves the efficiency and superiority of the proposed Gabor-SVM method.**

Keywords-gesture recognition; Gabor wavelet; SVM; PCA

I. Introduction

Hand gesture has become an important application in vision-based human-computer interfaces (HCIs) for the past decades because the traditional HCI input devices, such as mice and keyboard, cannot quickly respond to currently complicated interaction systems. For hearing impaired community, the development of automatic gesture translation based natural languages (e.g. the American Sign Language; ASL) is highly expected to improve their communication means among humans. For the recognition system of hand gestures based on images or videos, the posed-angle of gestures taking by a camera/webcam can usually be a critical factor in determining the effectiveness of the recognition systems.

Triesch and Malsburg [1] applied the method of elastic bunch-graph matching (EBGM) to the classification of hand postures for grayscale images. For EBGM method, the jets are described as vectors based on a 2D Gabor-wavelet transform. The results showed that their system can reach 86.2% recognition rates against complex background. This approach can achieve user-independent and scale-invariant recognition. However, this method is not view-independent and is computational inefficiency due to the matching process.

Chen and Tseng [2] proposed a multi-angle hand gesture recognition system for finger guessing games. To cope with various angles for hand gestures, they used three webcams set at front, left, and right directions of hand to capture gesture images. Then, three SVM classifiers are trained using the images acquired from the three cameras. After the training process, the constructed classifiers were fused by one voting and two plans of fusion to decide the gesture. The recognition rates of their system for the front, left, and right classifiers are 73.3%, 87.5%, and 92.5%, respectively. However, only 3 hand gestures were used in their work.

Amin and Yan [3] used PCA and Gabor filters to recognize the American Sign Language (ASL) finger alphabets from hand gesture images. The classification is then conducted with a method of fuzzy-c-mean clustering. The experimental results showed that the recognition rate of the ASL alphabets with average 93.23% accuracy can be achieved. However, the recognition rate of similar alphabets is relatively low in their approach.

In the past decade, the Gabor features have been successfully used in the fields of hand gesture and face recognitions [3,4]. However, Gabor features employed are of too high dimensionality to be used effectively. We proposed to deal with this problem by the PCA method to reduce the dimensionality of the feature space. The classification of hand gestures is then performed by the SVM method. Finally, a real-time video system on hand gesture recognition is presented.

The remainder of the paper is organized as follows: In Section 2, we describe how to construct the dataset of hand gestures and briefly outlines the methods of extracting the hand gesture features, followed by sketching the method of SVM. In Section 3, we present our experiment results. Finally, Section 4 gives a conclusion and some suggestions for future work.

II. System Description

This section consists of the major components such as image capturing and preprocessing, feature extraction and classification of hand gestures. A brief description is provided in the subsequent sections.

A. Image Capturing and Preprocessing

Some critical factors such as lighting condition, posed-angle and scale variability of hand gesture should be considered while collecting images for hand gesture recognition. The images of 11 hand gestures (see Fig. 1) for training and testing were collected with the same colored background by 10 signers, and each signer was requested to sign the same hand gesture 12 times; each time from a different angle and position. In order to verify the classification capability, the hand was cropped manually and resized to 20*20 pixels at the initial stage of the experiment. Fig. 2 shows that the pose of hand gestures with small angles

978-0-7695-3762-7/09 $26.00 © 2009 IEEE
DOI 10.1109/IIH-MSP.2009.96

(<45 deg.) and large angles (>45 deg.), which were taken six times separately to test the robustness of the proposed method. The dataset contains 120 (=10*12) images of each hand gesture and a total of 1320 (=10*12*11) images for 11 signs of hand gesture.

In a real time vision-based system, we first extracted the hands from a sequence of video images using the skin color information. Skin color may not be enough for tracking hands but it is often a fast convenient cue. We used the concept of "reference white" [5] as a method for lighting compensation and exploited the skin color model suggested by Soriano et al. [6] for converting a RGB color space into a normalized rgb color space. The segmented image of hand gesture was first resized to 20*20 pixels and then converted it into a grayscale image.

B. *Feature Extraction of Hand Gestures*

Features of hand gesture are collected by the following three steps. We first convolved the hand gesture images with the Gabor filters. The PCA method was then used to reduce the dimensionality of the Gabor-coded images. Finally, the Gabor-coded images were concatenated by the rows to form a discriminating feature vector. The details of the two methods are described in the following sections.

1) The Gabor filter: Gabor filters (wavelets, kernels) can capture the most significant visual properties such as spatial locality, orientation selectivity, and spatial frequency characteristics. Considering the preferable characteristics, we chose the Gabor features to represent the hand gesture images.

Mathematically, a 2D isotropic Gabor filter is the product of a 2D Gaussian and a complex exponential function. The general expression can be expressed as

$$g_{\theta,\gamma,\sigma}(x,y) = \exp\left(-\frac{x^2+y^2}{\sigma^2}\right)\exp\left(\frac{j\pi}{\lambda}\left(x\cos\theta + y\sin\theta\right)\right) \quad (1)$$

The parameter θ represents the orientation, λ is the wavelength, and σ indicates scale at orthogonal direction. However, with this set of parameters the Gabor filter does not scale uniformly as the parameter σ changes. It is better to use a parameter γ=λ/σ to replace λ so that a change in σ corresponds to a true scale change in the Gabor filter. Also, it is convenient to apply a 90^o counterclockwise rotation to (1), such that θ expresses the normal direction to the Gabor wavelet edges. Therefore, the Gabor filter can be alternatively defined as follows

$$g_{\theta,\gamma,\sigma}(x,y) = \exp\left(-\frac{x^2+y^2}{\sigma^2}\right)\exp\left(\frac{j\pi}{\gamma\sigma}\left(x\sin\theta - y\cos\theta\right)\right) \quad (2)$$

By selectively changing each of the parameters of the Gabor filter, one can tune the filter to a specific pattern arising in the image. Some examples of Gabor filter with different parameters (γ, θ, σ) are illustrated in Fig. 3.

By convolving a Gabor filter with image patterns, the similarity based on the Gabor response can be estimated. To emphasize three types of characteristics in images such as edge-oriented, texture-oriented, and a combination of both, one can change the parameters (σ, γ, θ) of the Gabor filter, where the variation of θ changes the sensitivity to edge and texture orientations, the variation of σ represents different "scales", and the variation of γ indicates the sensitivity to high/low frequencies.

A set of parameters of the Gabor filters used is γ=0.785, θ={0, π/2, 2π/5, π/4, π/5, 2π/11, π/7, π/8}, and σ={1, 2, 3, 4, 5}. Therefore, 40 Gabor responses from each image can be obtained. Each filter response is then converted into a pattern vector with 400 elements for a filter response of 20*20 pixels. By concatenating the 40 filter responses, the pattern vector has a dimensionality of 400*40=16,000. Fig. 4 shows the Gabor filter responses of hand gesture for sign "2".

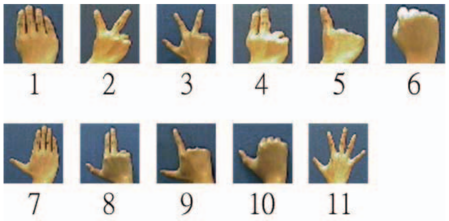

Figure 1. Hand gestures in the dataset

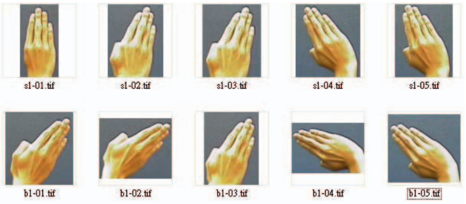

Figure 2. Hand gesture images with small angle (<45 deg.) in top row and with large angle (>45 deg.) in bottom row

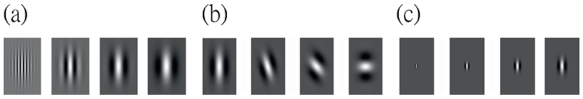

Figure 3. Examples of Gabor filters. Each example shows the real part of Gabor filter for different parameters. (a) γ={1/2, 3/2, 5/2, 7/2}; (b) θ={0, π/6, π/3, π/2}; (c) σ={4, 8, 12, 16}

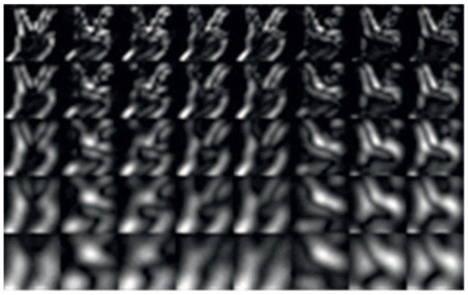

Figure 4. Gabor-coded image of hand-gesture "2"

2) Principal component analysis (PCA): The method of PCA [7] is a popular dimensionality reduction technique with the goal to find a set of orthonormal vectors in the data space, which can maximize the data's variance and map the data onto a lower dimensional subspace spanned by those vectors.

Consider a dataset with M images $x_i \in \Re^N \ (i = 1, \cdots, M)$ belonging to C subjects, and N is the number of pixels in the image. The total scatter matrix $S_T \in \Re^{N \times N}$ is defined as

$$S_T = \sum_{i=1}^{M} (x_i - \mu)(x_i - \mu)^T = AA^T \tag{3}$$

where μ is the global mean image of the training set, and $A = [x_1 - \mu \ \cdots \ x_M - \mu] \in \Re^{N \times M}$.

A direct computation of S_T is impractical due to the huge size $N \times N$ of the matrix. Instead of direct finding the eigenvector W_{PCA} of $S_{T,}$ we solve the eigenvalue problem, $RV_{PCA} = V_{PCA}\Lambda$, to obtain the eigenvectors, $V_{PCA} \in \Re^{M \times P}$, and the eigenvalues, $\Lambda = diag[\lambda_1 \ \cdots \ \lambda_P] \in \Re^{P \times P}$, with decreasing order $\lambda_1 \geq \cdots \geq \lambda_P > 0$, where λ_i is the nonzero eigenvalue of the matrix $R = A^T A \in \Re^{M \times M} \ (M \ll N)$. Then, the PCA subspace W_{PCA} is formed by multiplying the matrix A with the eigenvectors V_{PCA}, that is, $W_{PCA} = AV_{PCA} \in \Re^{N \times P}$. Therefore, the feature vector **y** of an image **x** is acquired by projecting **x** into the coordinate system defined by the PCA subspace, that is

$$y = W_{PCA}^T (x - \mu) \in \Re^P \tag{4}$$

C. *Classification of Hand Gestures*

In principle, one SVM classifier searches for an optimal hyperplane that maximizes the margins of their decision boundaries to ensure that their worst-case generalization errors are minimized, which is known as "structural risk minimization (SRM)."

To perform the classification between two classes, a nonlinear SVM classifier is applied by mapping the input data (x_i, y_i) into a higher dimensional feature space using a nonlinear operator $\Phi(x)$, where $x_i \in \Re^d$ and $y_i \in \{+1, -1\}$. Therefore, the optimal hyperplane can be computed as a decision surface

$$f(x) = \operatorname{sgn}\left(\sum_i y_i \alpha_i K(x_i, x) + b \right) \tag{5}$$

where $\operatorname{sgn}(\cdot)$ represents the sign function, and $K(x_i, x) = \Phi(x_i)^T \Phi(x)$ is the predefined kernel function that satisfies Mercer's condition [8]. In this research, the radial basis function (RBF) is used and it is defined as follows

$$K(x_i, x) = \exp\left(-\gamma \| x_i - x \|^2 \right), \ \gamma > 0 \tag{6}$$

where γ=0.25. The coefficients α_i and b in (5) can be determined by the following quadratic programming (QP) problem

$$\begin{aligned} &\max \left[\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right] \\ &s.t. \ \sum_i \alpha_i y_i = 0 \\ &0 < \alpha_i < C, \ \forall i \end{aligned} \tag{7}$$

The parameter C is a penalty that represents the tradeoff between minimizing the training set error and maximizing the margin, where $C = 8$ is determined empirically. Since the SVM is a binary classifier, it should be extended for an *m*-class problem in hand gesture recognition. We used the so called one against one approach, which is a pairwise method and needs to train $m(m-1)/2$ SVM classifiers. In addition, another two distance measures, i.e., Euclidean distance and Cosine similarity distance, are also computed to make a comparison with the SVM method.

III. Results and discussion

The dataset of hand gesture images was classified as a training set and a testing set, for which data set has 6*11*10=660 images (3 small angles and 3 large angles data selected randomly from each of the 10 signers for 11 hand gestures). The image of hand gesture was first convolved with Gabor filters to form a Gabor-coded image. The data of 40 filter responses were concatenated by the rows to form a pattern vector with a dimensionality of 16,000, which is further reduced by the PCA method to construct a discriminating feature vector. Finally, the classification of hand gestures was performed by SVM (C=8 and γ=0.25), a Euclidean distance, and a cosine similarity distance, respectively.

Fig. 5 shows the recognition rates for the three methods when different number of features acquired from the PCA method is used. The maximum recognition rates using SVM, the Euclidean distance, and the cosine distance are 95.2%, 93%, and 93%, respectively, with corresponding numbers of features being 100, 50, and 50, respectively. This result confirms the outstanding performance of the proposed Gabor-SVM method when compared to the other two methods. The analysis of the confusion matrix performed by the SVM method with a number of features of 100 (see Table 1) verifies the results and reveals which gestures are sources of errors.

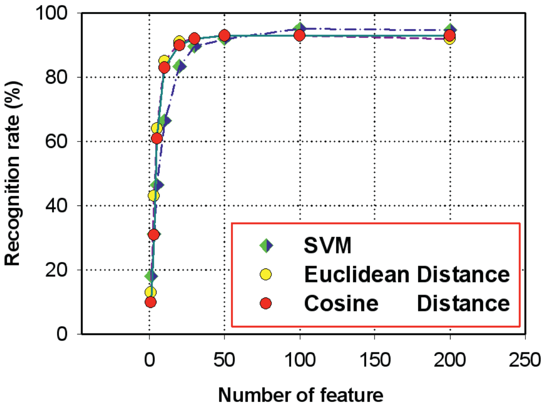

Figure 5. Results of recognition rate versus number of features used

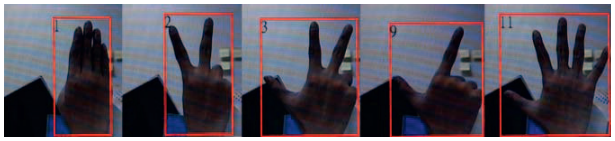

Figure 6. Some detection results of hand gestures for the proposed system

For instance, gestures 4 and 5 are very similar and thus are partly confused with the SVM method. The rate of gesture 5 misclassified to gesture 4 is 8.3%. However, gesture 8 is confused with gesture 9 having a misclassification rate of 6.6%. This is not really surprising; indeed gestures 8 and 9 have been selected to test the limits of recognition.

A vision-based hand gesture recognition system was carried out on a Pentium PC with a 3.4 GHz processor and 4GB DDR II memory. A sequence of video images were captured by a webcam, and then processed by skin-color segmentation to extract the hand from the image. The cropped image is further resized to 20*20 pixels. Some detection results of hand gestures are shown in Fig. 6. The processing rate of the proposed system is about 0.2 second for every frame, which has readily achieved the requirement of a real-time system.

IV. CONCLUSIN AND FUTURE WORK

We have proposed a novel Gabor-SVM method, which can achieve a recognition rate of 95.2% of hand gestures, and it is better than those of the other two methods, namely, Euclidean and cosine measures. The recognition results confirm the efficiency and superiority of the proposed Gabor-SVM method. Additionally, a hand gesture recognition system has been implemented with a processing rate of 0.2 second per frame. However, only 11 hand gestures are used with a limitation of wearing long sleeve clothes. In the future work, to increase the versatility of our hand gesture dataset, more different hand gestures need to be added and the limitation of wearing long sleeve clothes should be relaxed to accommodate the real environments.

ACKNOWLEDGMENT

This research was fully supported by a grant from National Science Council, Taiwan, under contract NSC-97-2221-E-212-035.

REFERENCES

[1] J. Triesch and C. von der Malsburg, “Robust classification of hand postures against complex backgrounds,” In: Proc. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition, Killington, Vermont, USA, Oct. 1996, pp. 170–175.

[2] Y. T. Chen, and K. T. Tseng, “Multiple-angle hand gesture recognition by fusing SVM classifiers,” In: IEEE conference on Automation Science and Engineering, Scottsdale, AZ, USA, Sep. 2007, pp. 527-530.

[3] M. A. Amin, and H. Yan, “Sign language finger alphabet recognition from Gabor-PCA representation of hand gestures,” In: Proc. of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, August 2007, pp. 2218-2223.

[4] C. Liu, “Gabor-based kernel PCA with fractional power polynomial models for face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, pp. 572–581, May, 2004,.

[5] R. L. Hsu, A. M. Mohamed, and A. K. Jain, “Face detection in color images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, pp. 696-706, May, 2002,.

[6] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, “Skin detection in video under changing illumination conditions,” In: Proc. of 15th International Conference on Pattern Recognition, Barcelona, Spain, vol. 1, 2000, pp. 839-842.

[7] M. Turk, and A. Pentland, “Eigenfaces for recognition,” J. Cogn. Neurosci., vol. 3, pp. 71-86, January, 1991.

[8] V. N. Vapnik, Statistical learning theory, John Wiely and Sons, New York, 1998, pp. 423-424.

TABLE I. CONFUSION MATRIX FOR THE RECOGNITION RESULTS BY SVM WITH A NUMBER OF FEATURES OF 100.

	1	2	3	4	5	6	7	8	9	10	11
1	93.3	0	0	1.6	0	1.6	1.6	0	0	0	1.6
2	0	100	0	0	0	0	0	0	0	0	0
3	0	1.6	98.3	0	0	0	0	0	0	0	0
4	1.6	0	0	96.7	0	0	1.6	0	0	0	0
5	0	1.6	0	8.3	88.3	0	0	0	1.6	0	0
6	0	0	0	0	0	100	0	0	0	0	0
7	1.6	0	0	0	0	0	88.3	6.6	0	0	3.3
8	0	0	0	0	0	0	1.6	90.0	6.6	1.6	0
9	0	0	1.6	0	1.6	0	1.6	1.6	91.6	1.6	0
10	0	0	0	0	0	0	0	0	0	100	0
11	0	0	0	0	0	0	0	0	0	0	100