The Study of Gesture Recognition Based on SVM with LBP and PCA

Jingzhong Wang[1], Xiaoqing Xu[1,2], and Meng Li[1,2]
[1]North China University of Technology, Beijing, China
[2]Beijing Electronic Science and Technology Institute Beijing, Beijing, China
Email: jingzhongwang@163.com, xxq804@besti.edu.cn, 410068528@qq.com

***Abstract*—This paper gives an improved gesture recognition algorithm. First, obtain the complete hand-type region of the image, using the image preprocess algorithm such as YUV color segmentation, image differencing, connected domain detection. Then process images through contour detection, have the feature extraction and compression by LBP transform and principal component analysis. Finally, use support vector machine as a training machine learning algorithms to build classifier classification. To research a total of 630 gesture images, the experimental results show that the proposed method for gesture recognition, which can improve the recognition rate and speed effectively, and the recognition rate reaches 94.22%.**

***Index Terms*—gesture recognition, LBP, PCA, SVM, machine learning**

I. Introduction

Sign language recognition is a current popular and comprehensive research in pattern recognition and machine learning. In the field of human-computer interaction, in order to enable users to obtain better operating experience, vision-based sign language recognition research has become the core gradually [1].

The core technology of pattern recognition is feature extraction and classification. Common methods of feature extraction can use texture feature such as LBP, which has been applied in face recognition [2]; SIFT (Scale Invariant Feature Transform) feature has maintain invariance to rotate, scale zoom and brightness variation. However, there exist a number of shortcomings, such as fewer number of successful match point, slower, specific application purposes and background; SURF, the improved algorithm of SIFT, reduces the complexity of the algorithm and improves computing speed has greatly, but the effectiveness has declined [3]; Hu moment is a common characteristic according to the geometric characteristics with simple calculation process, but it has no obvious advantage to recognition rate [4]. Choose a high-dimensional vector feature makes calculation complex and a longer follow-up time. Therefore, in order to make the process simpler and efficiently, can remove the data which has smaller effect on the classification, for example principal component analysis [5]. Then the classification ability can be improved effectively.

Manuscript received January 17, 2015; revised June 12, 2015.

SVM algorithm works in pattern recognition, regression estimation, probability density function estimation and other aspects. The accuracy of the algorithm higher than traditional learning algorithms in some fields, it has a better learning performance [6]. In PCA algorithm, vectors between each component are orthogonal, so there is independence between the same sample characteristics. Calculate the similarity independently for each sub-vector in SVM training process, and find an optimal hyperplanes, just can make up for the defect mentioned above.

This paper researches phonetic alphabet gesture recognition based on visual. It combines with a variety of methods to detect gestures like YUV color detection [7] and image differencing, then takes feature extraction by LBP transforming and dimension reduction of PCA, applies SVM classifier to train and learn the classification, finally gets recognition results.

II. Key Technology

A. Local Binary Pattern (LBP)

Local Binary Pattern (LBP) is a very effective method of feature extraction with rotation invariance. LBP is an algorithm which makes the binary encoding about each pixel (x_c, y_c) in the image, and converts it to a decimal number according to a certain manner [8].

In order to adapt to different scales of LBP texture features description, the typical operator 3*3 neighborhood model extended to an arbitrary circular neighborhood, and allows for any number of sampling points in the radius R of the circular neighborhood. When the sampling point does not fall within the pixel center position, its gray values calculated using bilinear interpolation. A circular neighborhood of radius R which has P pixels is indicated by symbol LBP (P, R). The common LBP operators have LBP (8, 1), LBP (16, 2), LBP (8, 2), etc.

Ojala [9] proposed a uniform pattern to reduce the dimension for the model type of LBP operator which can solve the problem about too many binary modes and improve the statistical. In an actual image, the vast majority of LBP models contain only twice transitions at most from 1 to 0 or from 0 to 1. “uniform pattern” shall: When a circulating binary number which the LBP corresponding to up to two transitions such as from 0 to 1

doi: 10.18178/joig.3.1.16-19

or from 1 to 0 , the binary pattern of LBP corresponding to is called an uniform pattern class .

With this improvement, the type of binary mode is reduced greatly, without losing any information. The number of modes was reduced from 2^P to P (P-1)+2, where P represents the neighborhood sampling points within a set. That makes the feature vector less dimension, and can reduce the influence caused by high frequency noise.

B. Principal Components Analysis (PCA)

Principal Components Analysis is the projection of high-dimensional data to lower-dimensional space, compress the size of the original data matrix, while retaining the original data information largely, and achieve the purpose of dimensionality reduction. Therefore, PCA method is used widely in machine learning, data classification and pattern recognition. The basic steps are as follows:

1) Assuming there are m data samples $\{x_m\}$ and each data is n-dimensional, then construct a sample matrix M size of m * n.
2) Averaging row of the matrix M, the formula of mean vector $\bar{x}$ about data in m samples as follows:

$$\bar{x} = \frac{1}{M}\sum_{m=1}^{M} x_m \tag{1}$$

3) Calculate covariance matrix S of the sample matrix , the formula as follow:

$$S = \frac{1}{M}\sum_{m=1}^{M}(x_m - \bar{x})(x_m - \bar{x})^T \tag{2}$$

4) Do singular value decomposition (SVD) to the covariance matrix, and obtain the eigenvalues and eigenvectors.
5) Make the eigenvalues in descending order, retain its main components (the first K eigenvectors) of the sample which used to represent.
6) To convert the data into a new space which from the K eigenvectors constructing [10].

C. Support Vector Machine (SVM）

Support Vector Machine was a new general learning method by Vapnik *et al.* firstly. It is based on the basis of statistical theory and structural risk minimization principle, can solve some practical problems such as the small sample, nonlinearity, high dimension, local minima points and so on. Support Vector is a set of feature subset for the training set, so that the linear partition for the feature subset is equivalent to divide the entire data set [11].

As shown in the picture, in order to apart the white and black points, we can draw a straight line countless. We call these lines hyperplane, choose one of them, and it meets $H:(\omega x)+b=0$. In the picture, H1 or H2 is the closest to the hyperplane from all types of samples respectively and they parallel to the plane of hyperplane. Distance between H_1 and H_2 is called classification interval. Make the black points = -1, white points = +1, can write the H_1 and H_2 expression are $\omega x_i + b = -1, \omega x_i + b = 1$. According to the distance formula between the planes, easy to obtain the distance between H_1 and H_2 to H is $1/\|\omega\|$, classification interval is $2/\|\omega\|$. Provides that the surface which make classification interval $2/\|\omega\|$ largest is the optimal hyperplane. See Fig. 1, the training sample points crossed by the H1 and H2 constitute a support vector.

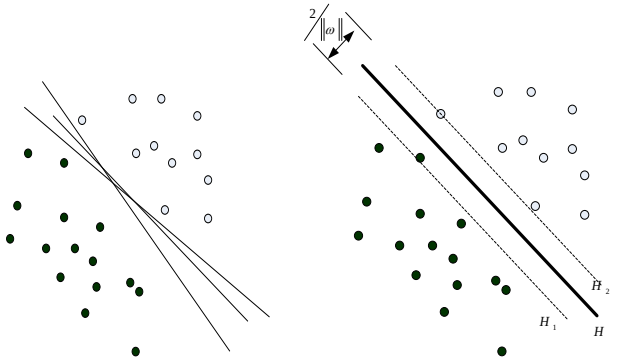

Figure 1. SVM schematic diagram.

In the case of linear inseparable, one hand can still divide them with a straight line, which introduced the concept of penalty function; on the other hand can make the input vector map to a high dimensional feature space, then construct optimal hyperplane in this space.

III. Gesture Recognition System

The process framework of gesture recognition system implementation is showed in Fig. 2:

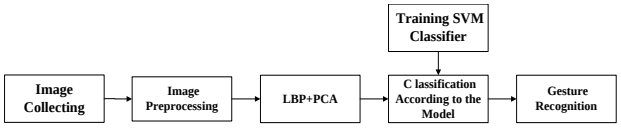

Figure 2. Gesture recognition system flowchart

See Fig. 2, the design of sign language recognition system is divided into two parts.

For the first part: create the gesture samples, preprocessing the image and take the feature extraction of the sample set, and finally training a classifier.

The second part is built on the basis of the first part, for a single gesture image. Gesture image through the same pre-processing and feature extraction process, and then make the resulting gesture feature vectors classified by classifier to obtain the recognition result.

A. Image Pre-Processing

The purpose of gesture image preprocessing is to detect hand position of collected gesture images from the camera, remove the non-target area to obtain the gesture area image. See Fig. 3 which about the implementation of image preprocessing framework.

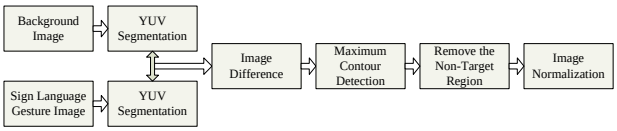

Figure 3. Gesture detection system flowchart

Through image preprocessing, can obtain the hand type binary image after be normalized size of 32*32. By combining with YUV color detection [12] and the background subtraction method, reduce the impact of the background color to the gesture detection. After detecting the maximum connected domain, using the improved projection method to remove the arm and other non-target area which based on the direction of the gesture, so that obtain accurate hand-type area for the subsequent feature extraction and recognition.

B. Feature Extraction

LBP can be applied to face recognition or texture detection and so on. In face recognition, common method is that use LBP for face's gray scale image and obtain feature image after the transformation. However, if selected the hand type gray scale image to LBP, it is not suitable as a gesture classification characteristic.

Contour is one of the important features of an image. Compare with the edge of graphics, outline is a hollow graphic which without interior points. For example, in a picture has 3*3 rectangular points, and then remove the middle point of that, the rest is the contour. In terms of hand-shaped contour image, after LBP transformation, can remove part of the information which contribution rate is less, so that reduce the dimension of feature vectors.

Characteristics of this paper are based on gesture outline image. After equivalence mode LBP conversion, anti-color image processing, morphological dilation operation to obtain the feature image, and then through the PCA dimension reduction obtain the feature vectors. Proceed as follows:

Step 1: Contour extraction, LBP transform in equivalent model. See Fig. 4 which is part of LBP converted images.

Figure 4. Part gesture images after LBP processing

Step 2: Take principal component analysis and obtain the feature vector after dimensionality reduction.

C. Classification and Recognition

Make the sample feature vector after the feature extraction to machine learning, and train SVM classifier to obtain a model.

In the gesture recognition system, the test image through preprocessing, feature extraction, projecting it onto the PCA mapping matrix to obtain the feature vector, then have a prediction based on classifier training model, the result is the classification results.

IV. Experimental Results

A. Gesture Image Recognize

In the gesture recognition system, according to 30 Pinyin alphabet gestures as a reference, collect three sets of gesture (three different people's hands respectively) to form a gesture library. Each gesture of every set has recorded seven letters, so that there are totally 630 gesture maps in three gestures library. Each image's size is 640*480; the format is a bitmap (BMP).

According to this algorithm, have a lot of experiments to test their impact on the recognition rate for static gestures. Choose two images from each class in every gesture library, and obtain 180 images as samples. While the remaining 450 images as test images. According to the PCA algorithm, select different value of the main component K to identify the effects of the recognition.

The images recognition results of gesture library as shown in the Table I below:

TABLE I. Sign Language Recognition Results of the Test Set

Value of K	Average Rec-rate (%)
50	94.22
40	93.78
30	93.33
20	93.11
10	90.22

The experimental results show that the algorithm can identify the alphabet gestures effectively. When K is equal to 50, the average recognition rate reaches 94.22, and can complete the identification of 150 images in 43 seconds on average.

When the value K becomes smaller gradually, the recognition speed can increase, however, the recognition rate has decreased. Therefore, when choosing the value of K, should weigh the speed and accuracy. When K is equal to 30, the recognition rate is good with less time, so the gesture recognition system can select this value for K.

The method mentioned in ref. [13] that using SURF feature package and SVM to classify gestures, the average recognition time of each image is 0.761/S; combined with Hu moments and the BoF-SURF algorithm, the average recognition time of each image is 0.838/S. In contrast, the proposed algorithm has improved the recognition speed significantly, and the recognition rate is better.

B. Gesture Video Recognize

As the same way to a simple gesture video, also can recognize its meaning. In another word, change the video to pictures, then can use same algorithm to recognize a video meaning by key frame. The steps as follow:

Step 1: The overall time is divided into four average parts and then can obtain five images from a video, see Fig. 5:

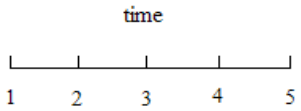

Figure 5. Key frame based on time

Such as "hello" in Chinese gesture like that, see Fig. 6:

Figure 6. Five key frame of hello

Step 2: Take the same image pre-process and feature extraction algorithm for each gesture image. The results see Fig. 7.

Figure 7. Five key frame after feature extraction

Step 3: Calculate the Euclidean distance of feature vector to sample vectors respectively. And can obtain five distances for each sample.

Step 4: Weighted sum the five distances, then consider the result is the sample's meaning which the minimum value corresponding to.

There are total seven kinds of gesture videos for my test. The recognize rate can reach more than 80%. The experimental results show that the way is effective.

V. Conclusion

This paper researches the recognition of Chinese phonetic alphabet 30 gestures, give the algorithm that LBP combined with PCA in conjunction with the machine learning to recognize sign language. First take gesture detection, then feature extraction based on the uniform pattern of LBP and principal component analysis, and finally use LIBSVM [14] multi-classifier for classification. Experiments show that the method can detect and recognize gesture efficiently with the stability to image panning and zooming as well as light.

Acknowledgment

This work was supported in part by China the National Natural Science Funds under Grant No.61371142 and Beijing innovation team building promotion plan under Grant ID HT20130502.

References

[1] Z. Q. Feng and Y. Jiang, "A survey of hand gesture recognition," *Journal of University of Jinan*, vol. 4, 2013.

[2] W. Wang, *et al.*, "Face description and recognition by LBP pyramid," *Journal of Computer-Aided Design & Computer Graphics*, vol. 1, pp. 94-100, 2009.

[3] Q. Y. Zhang, *et al.*, "Hand gesture recognition based on bag of features and support vector machine," *Journal of Computer Applications*, vol. 32, no. 12, pp. 3392-3396, 2012.

[4] Q. Y. He, Y. Ge, and L. Q. Wang, "Gesture recognition algorithm based on invariant moment and edge detection," *Computer Engineering*, vol. 31, no. 15, pp. 165-166, 2005.

[5] R. J. Li, Q. L. Han, and X. H. Yang, "New optimization method of PCA face recognition," *Journal of Dalian Jiao Tong University*, vol. 29, no. 4, pp. 48-51, 2008.

[6] Q. Yang and J. Y. Peng, "Sign language recognition based on a multifeatures SCMs classifier," *Electronic Engineering & Product World*, vol. 4, pp. 39-43, 2009.

[7] L. M. Tao, Z. Y. Peng, and G. Y. Xu, "The features of skin color," *Journal of Software*, vol. 12, no. 7, pp. 1032-1041, 2001.

[8] M. Topi, O. Timo, P. Matti, *et al.*, "Robust texture classification by subsets of local binary patterns," in *Proc. 15th IEEE International Conference on Pattern Recognition*, 2000, vol. 3, pp. 935-938.

[9] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.

[10] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433-459, 2010.

[11] T. Joachims, "Making large scale SVM learning practical," in *Advances in Kernel Methods – Support Vector Learning*, MIT Press, 1999.

[12] D. Qiu, "Research about skin color detection base on HSV and YCrCb color space," *Computer Programming Skills & Maintenance*, vol. 10, pp. 74-75, 2012.

[13] Y. H. Sui and Y. S. Guo, "Hand gesture recognition based on combining Hu moments and BoF-SURF support vector machine," *Application Research of Computers*, vol. 31, no. 3, pp. 953-956, 2014.

[14] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.

J. Z. Wang was with Inner Mongolia University, Inner Mongolia, China. He is now working in the College of Information Engineering, North China University of Technology, Beijing, 100144China (Corresponding author, phone: 13001204464 e-mail: jingzhongwang@163.com).

X. Q. Xu is now with the Department of Computer Science, Adjunct Professor of Computer Science Electronic Science and Technology Institute Beijing, Beijing, and zip code 100070 China.(phone:86-10-63740055-2106. e_mail:xxq804@besti.edu.cn).

M. Li is now with the College of Information Engineering, North China University of Technology, Beijing, and zip code 100144 China.(Corresponding author, phone: 135-110-58776; e-mail: 410068528@ qq.com).