RECOGNITION OF HAND GESTURE BASED ON GAUSSIAN MIXTURE MODEL

Jia Jia, Jianmin Jiang, Dong Wang

EIMC, School of Informatics, University of Bradford, West Yorkshire BD7 1DP, UK
jjia@bradford.ac.uk, j.jiang1@bradford.ac.uk, D.Wang6@brad.ac.uk

ABSTRACT

This paper presents a new method for gesture recognition of Human beings' hand. This method integrates the features of shape, color and orientation histograms, which are extracted from images, and estimate the comparability with all the different types of gestures by a proposed Expectation-Maximization algorithm in Gaussian Mixture Model. The classification results were presented based on the values of likelihood compared with all the types of pre-assigned images, and the performance of this approach in an experiment is shown that the proposed method works well.

1. INTRODUCTION

Human gesture has its specific meanings and is widely used for communications between deaf people. It is considered as a very important function in many practical communication applications. Recently, hand gesture recognition has gained a lot of interests, which plays a crucial role in a wide range of applications, such as automatic sign language understanding, entertainment, and human computer interaction (HCI). Because hand gestures are natural and intuitive in providing rich information to computers without extra cumbersome devices, they can offer a great potential for next generation user interfaces, being especially suitable for large scale displays, 3D volumetric displays or wearable devices.

In addition, there has recently been a growing interest in gesture-recognition systems by a number of researchers, providing some novel approaches since early nineties. Chaitanya Gurrapu [1] adopts GMM to classify the human body's gestures and track its changes through time using HMM. The key features extracted from body images are polygonal vertices obtained from body shapes. Accordingly, GMM is trained on the vertices feature and the relationship between vertices which is represented in the form of gradient of the line joining two vertices. The final accuracy is close to 98%. Yang Liu [2] used a method integrating shape and depth information for robust hand tracking. Shape is the primary measurement which builds an important function describing areas of state-space and contains critical information about the posterior. Recognition rate is between 70% and 92% under various numbers of samples. Sebastien Marcel [3] presented a hand gesture recognition algorithm based on input/output Hidden Markov Models. This approach achieves a recognition rate between 90% and 100% of the sequence.

In our system, we addressed the core issue of gesture recognition in extracting robust features, leading to a more accurate estimation. The new approach we propose is different from existing efforts reported in the literature. It focuses on estimating the gesture contained in an image by analyzing different complex features including shape, color and orientation histogram quantized in Gaussian Mixture Model (GMM).

GMM is a widely used statistical model in many applications of pattern recognition, which is often regarded as a versatile modeling tool as it can be used to approximate any probability density function (PDF) given a sufficient number of components, and impose only minimal assumptions about the modeled random variables. The advantage is including a rigorous statistical basis, the possibility of encoding spatial, color, texture and motion features in a unified system, and the ability to trade off accuracy of representation against data volume. Due to such advantages, our proposed technique builds upon the GMM to estimate the mutative meaning of human gestures in a compact and precise manner.

The rest of this paper is organized as follows. An overview of principle of Gaussian Mixture Model is described in Section 2. The method of training and parameter estimation is covered in Section 3. While Section 4 is dedicated to description of how images are tested, Section 5 is devoted to the experiments. Finally, concluding remarks and further area of research are provided in Section 6.

2. DESCRIPTION OF PRINCIPLE OF GAUSSIAN MIXTURE MODEL (GMM)

GMM is one of the most widely used mixture modeling techniques. It's a simple model and is reasonably accurate when data are generated from a set of Gaussian distributions [4, 5]. Let $X_i = \{x_t, 1 \le t \le T^i\}$, denote the feature vectors for data points from the *i*-th class. They are modeled by a total number of J Gaussians as follows:

$$P\left(X_i \middle| \theta_{GMM}^i\right) = \prod_{t=1}^{T^i} \sum_{j=1}^{J} P(z_j) P_{z_j}\left(x_t \middle| u_j, \Sigma_j\right) \tag{1}$$

Where θ_{GMM}^i includes all the model parameters, i.e. $\{P(z_j), \mu_j, \Sigma_j, 1 \le j \le J\}$. $P_{z_j}\left(x_t \middle| \mu_j, \Sigma_j\right)$ is the Gaussian distribution for the *j*-th class, with a mean vector μ_j and a covariance matrix $\sum_j$ as:

$$P_{z_j}\left(x_t \middle| \mu_j, \Sigma_j\right) = \frac{1}{(2\pi)^{D/2} \left|\Sigma_j\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(x_t - \mu_j\right)^T \sum\nolimits_j^{-1} \left(x_t - \mu_j\right)\right\} \tag{2}$$

where D is the dimension of the feature vector x_t. Usually, Σ_j is set to be a diagonal matrix as $diag\{\sigma_{jd}^2 : 1 \le d \le D\}$ in order to reduce the size of parameter space.

It can be seen from Equation (1) that the data points of a specific class are generated from multiple Gaussian models with an identical weight $P(z_j)$. We define $\omega_j = P(z_j)$.

In other words, an integrated Gaussian mixture model contains three basic parameters: Mixture weight, Mean vector and Covariance matrix, which can be represented as:

$$\lambda = \{\omega_j, \mu_j, \Sigma_j\} \tag{3}$$

where ω_j is the mixture weight, μ_j is the mean vector, and Σ_j is the covariance matrix. We use λ to stand for every single image. Additionally, we use $b_j(x) = P_{z_j}\left(x_t \middle| \mu_j, \Sigma_j\right)$ and $\sum_{j=1}^{J} \omega_j = 1$.

3. TRAINING METHODS AND PARAMETER ESTIMATION

For training purposes, our primary work is to find the parameter λ which can stand for the feature vector of every certain image. A normal method is the maximum likelihood (ML) estimation via expectation-maximization algorithm [4, 6, 7]. The ML means attempting to find the certain λ from the image which is used for training purposes in order to get the maximum likelihood.

For example, we extract the feature vectors $X = \{x_1, \cdots, x_T\}$ from an image by selecting the feature where the T is the number of features and the likelihood of GMM is defined as:

$$P\left(X \middle| \lambda\right) = \prod_{t=1}^{T} p\left(x_t \middle| \lambda\right), \tag{4}$$

According to the fact that the $P\left(X \middle| \lambda\right)$ is nonlinear function, we should use the way of ML to estimate the parameter of GMM until the $P\left(X \middle| \lambda\right)$ is convergent.

The method of algorithm estimation starts from an initial guess λ for the new model parameters $\overline{\lambda}$ to be estimated, in order to get a relationship of $P\left(X \middle| \overline{\lambda}\right) \ge P\left(X \middle| \lambda\right)$. Then transform the $\overline{\lambda}$ into the initial model parameter λ. This step will be repeated until $P\left(X \middle| \lambda\right)$ is convergent. During the iteration, the following estimation ensures that the approximation of GMM is achieved via the nature of monotonic increase:

$$p\left(i \middle| x_i, \lambda\right) = \frac{\omega_i b_i(x_t)}{\sum_{k=1}^{T} \omega_k b_k(x_k)} \tag{5}$$

Where the estimation of mixture weight is

$$\omega_i = \frac{1}{T} \sum_{t=1}^{T} p\left(i \middle| x_t, \lambda\right) \tag{6}$$

The estimation of mean vector is:

$$\mu_i = \frac{\sum_{t=1}^{T} p\left(i \middle| x_t, \lambda\right) x_t}{\sum_{t=1}^{T} p\left(i \middle| x_t, \lambda\right)} \tag{7}$$

The estimation of covariance is:

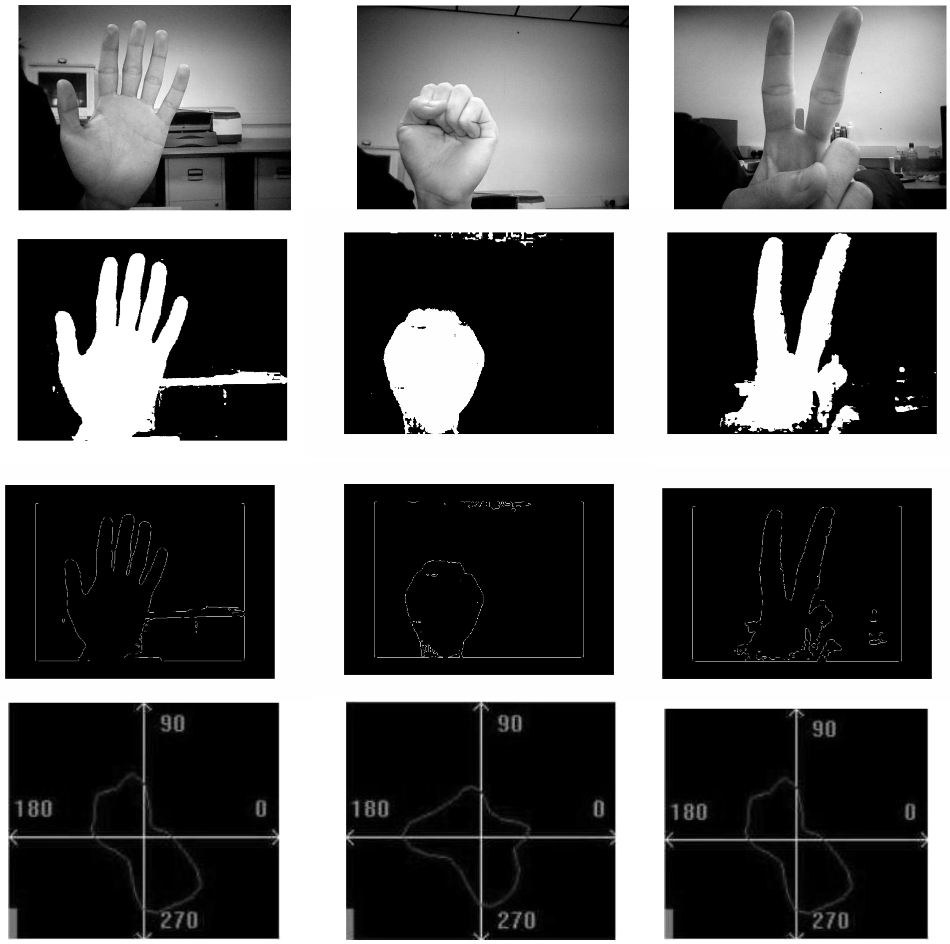
$$\Sigma_i^2 = \frac{\sum_{t=1}^{T} p\left(i \middle| x_t, \lambda\right) x_t^2}{\sum_{t=1}^{T} p\left(i \middle| x_t, \lambda\right)} - \mu^2 \tag{8}$$

4. METHODS OF TESTING

We use the maximum of a posterior criterion to differentiate all images, which means that the likelihood between testing images and pre-assigned images of each different type are calculated in order to compare the results and select the maximum numerical value. Accordingly, the testing image is ranged to a certain type, in which it has the maximum numerical value of likelihood compared with other images. In this way, we can use the equation below to describe the proposed process:

$$\hat{S} = \arg\max_{1 \le k \le S} \Pr(\lambda_k | X) \tag{9}$$

where S is the total of all pre-assigned different types, $\hat{S}$ is the certain type which the testing image is classified to, λ_k is the model of pre-assigned type *K*, and *X* is the vector of features of the testing image.



(a)
(b)
(c)
(d)

Fig.1. Classification of Hand Gesture: (a) Three basic classes of hand gesture; (b) colour extraction of each class of hand gesture; (c) shape extraction of each class of hand gesture; (d) orientation histogram of each class of hand gesture

This equation can be transformed to another by Bayesian rule:

$$\hat{S} = \arg\max_{1 \le k \le S} \frac{p(X|\lambda_k)\Pr(\lambda_k)}{P(X)} \tag{10}$$

for $\Pr(\lambda_k) = 1/S$, we have: $\hat{S} = \arg\max_{1 \le k \le S} P(X|\lambda_k)$. By calculating their logarithms, we have:

$$\hat{S} = \arg\max_{1 \le k \le S} \sum_{t=1}^{T} \log p(x_t|\lambda_k) \tag{11}$$

5. EXPERIMENTS

To evaluate the proposed algorithm, we collected a group of Hand-gestures as the test data set, and then classified all of them into 3 basic classes as shown in Fig. 1 (a). The first one is a hand with all fingers outstretched. The second one is considered as a fist, and the third one involves only 2 outstretched fingers (forefinger and middle finger) symbolizing a victory.

Following the described algorithm, three significant features are extracted, which include color, shape, and orientation histogram from all the image. The feature of color is extracted in the YCbCr Color Space. Fig.1 (b) shows the results of the color extraction of every image in Fig.1. The feature of shape is extracted by Canny Edge Detector. The results are shown in Fig.1 (c). Orientation histogram [8, 9] is one of useful features which offers robustness to lighting changes and give translational invariance. The orientation histograms for each gesture are illustrated in Fig.1 (d).

We evaluate the performance by using a total number of 450 gesture images, which were derived from the "Sebastien Marcel Static Hand Posture Database" [10]. As these images are captured against different backgrounds, it would help to test the robustness of the proposed algorithm if the order of images is randomized in the data set. Trained by 8 images from the database for each gesture, the proposed technique is evaluated by extensive experiments and their results are measured by error rate, which are summarized in Table-I:

Table-I: Summary of Experimental Results

Total Number	450
Images for Training	24
Amount of testing images	426
Error rate (%)	5.37%

6. CONCLUSION AND SUMMARY

In this paper, we described a proposed algorithm for human gesture recognition and demonstrated its discriminative ability for recognition of gestures on a large database of images. By using Gaussian Mixture Model, we have shown that multiple features extracted from gesture images could be organized and controlled by GMM to formulate new discriminating vector for classification and recognition of human gestures. The application of Gaussian Mixture Model illustrates the advantage that it provides improved performance over other existing methods, yet requiring only modest computational cost to complete the gesture recognition. Further research can be identified to focus on the issue of extendibility and selection of primary features as such that other pattern recognitions can be achieved, especially inside digital videos.

7. ACKNOWLESGMENT

The GMM code used here is based in part on Netlab Toolbox, and the work is part of the LIVE project initiatives funded under FP6 programme.

8. REFERENCES

[1] Chaitanya Gurrapu and Vinod Chandran, "Gesture Classification Using A GMM Front End and Hidden Markov Models", Proceedings of the 3rd IASTED International Conference on Visualization, Imaging, And Image Processing, Spain, pp. 609-612, September, 2003.

[2] Yang Liu and Yunde Jia, "A Robust Hand Tracking and Gesture Recognition Method for Wearable Visual Interfaces and Its Applications", Proceedings of the Third International Conference on Image and Graphics, IEEE Computer Society, USA, pp. 472-475, Dec 2004.

[3] Sebastien Marcel, Oliver Bernier, Jean-Emmanuel Viallet and Daniel Collobert, "Hand Gesture Recognition using Input-Output Hidden Markov Models", Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, IEEE Computer Society, USA, pp. 456-461, 2000.

[4] C.M.Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[5] R.O.Duda, P.E.Hart, and D.G.Stork, *Pattern Classification*, volume 2, Wiley-Interscience Publication, 2000.

[6] A.P.Dempster, N.M.Laird and D.B.Rubin, "Maximum likelihood from incomplete data via EM algorithm", Journal of the Royal Statistical Society, Series B (Methodological), Vol. 39, No. 1, pp. 1-38, 1977.

[7] J.Bilmes, "A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models", Technical Report ICSI-TR-97-021, University of California Berkeley, 1998.

[8] William T. Freeman and Michal Roth, "Orientation Histograms for Hand Gesture Recognition", IEEE International Workshop on Automatic Face and Gesture Recognition, IEEE Computer Society, Switzerland, pp. 296-301, June 1995.

[9] Hyung-Ji Lee and Jae-Ho Chung, "Hand gesture recognition using orientation histogram", Tencon 99. Proceedings of the IEEE Region 10 Conference, South Korea, pp. 1355-1358 Vol. 2, Dec 1999.

[10] S. Marcel, "Hand posture recognition in a body-face centered space", Proceedings of the Conference on Human Factors in Computer Systems, ACM, USA, pp. 302-303, 1999.