



Research on Dynamic Facial Expressions Recognition

Xiaoning Peng & Beiji Zou

School of Information Science and Engineering

Central South University

Changsha 410083, China

E-mail: hhp xn@mail.csu.edu.cn

Department of Computer Science and Technology of Huaihua College

Huaihua 418008, China

E-mail: hhp xn@163.com

Lijun Tang

School of Information Science and Engineering

Central South University

Changsha 410083, China

E-mail: hhp xn@mail.csu.edu.cn

Ping Luo

Department of Computer Science and Technology

Tsinghua University

Beijing 100084, China

E-mail: luop@mail.tsinghua.edu.cn

This work is supported in parts by the National Natural Science Foundation of China under Grant No.60673093 and supported by Hunan Provincial Natural Science Foundation of China under Grant No.06JJ2065 and supported by Program for Changjiang Scholars and Innovative Research Team in University of China.

Abstract

Human-computer intelligent interaction (HCII) is usually based on facial expression recognition. A dynamic facial expression recognition method based on video sequence is proposed in this paper, which uses Gaussian of Mixture Hidden Markov Model. Firstly, we get some special facial expression regions, in which the motion features are extracted and described as phase form and then constituted to eigen-sequences. Secondly we use Gaussian of Mixture Hidden Markov Model to learn and test these eigen-sequences, and recognize six universal facial expressions: angry, disgust, fear, happy, sad and surprise. And we developed an experimental system based on our algorithm. The experimental results show that the computing time and the error of vector quantization is reduced, while the classification efficiency is improved.

Keywords: Facial expressions recognition, Optical flow, Gaussian of Mixture Hidden Markov Model, Eigen-sequences

1. Introduction

Human-computer intelligent interaction (HCII) is an emerging field of science aimed at providing natural ways for humans to use computers as aids. To interact with the human, the computer must be equipped with human communication skills, such as understanding, distinguishing and identifying the capacity of human emotional state. Facial expression includes rich information about human emotions, it is the main carriers of emotions, and it is an important way of understanding emotions. Facial expression plays a very important role in human life, and it is a main way of human nonverbal communication. In recent years, the reasons for renewed interest in facial expression recognition are multiple, but mainly due to people have more interests about Human-computer intelligent interaction.

Facial expression recognition deals with the classification of facial motion and facial feature deformation into abstract classes that are purely based on visual information(B. Fasel, J. Luettin, 2003). The expression recognition method has very much, according to expression data these approaches can be divided into two main categories: one kind is based on

the static image method, another kind is based on the dynamic image sequence method. In static approaches recognition of facial expression is performed using a single face image. These approaches are, mainly, based on the Facial Action Coding System (FACS)(Ekman P, Friesen WV, 1978), a presentation developed in order to allow human psychologists to code expressions from still images. Such works fail to incorporate the timing of expressions, which is a critical parameter in emotion recognition. Mase(K.Mase, 1991) used optical flow to estimate facial muscle action and obtained 80% recognition rate in user-trained mode, which also belongs to this approach. But the temporal characteristics of facial expressions were not fully utilized in Mase's method. In order to manifest motion information between frames in facial expression changing, another approach is produced, which is based on dynamic image sequence. J.Lien(J. Lien, 1998) used this method, he analyzed facial expression and estimated expression intensity based on dense flow and Hidden Markov Model (HMM). But because he used discrete HMM, much important information was lost in quantifying vector process, complex computation was needed in determining quantification code process.

In view of insufficiency in the expression recognition based on the static image method, this article uses the dynamic facial expression recognition method, which will fully use the temporal and spatial information when expression changes. We get some special facial expression regions, in which the motion features are extracted and described as phase form and then constituted to eigen-sequences. In view of insufficiency in vector quantization based on discrete HMM, this paper describes facial expression time-based eigen-sequences with 1-st order left-right Gaussian of Mixture Hidden Markov Model (MHMM), and recognize six universal facial expressions: angry, disgust, fear, happy, sad and surprise. Some examples is as follow(Neutral Anger Disgust Fear Happiness Sadness Surprise

Figure 1):

2. The Basic Structure of Facial Expression Recognition Experimental System

Facial expression recognition deals with the classification of facial motion and facial feature deformation into abstract classes that are purely based on visual information. Our basic structure of facial expression recognition experimental system consists of three steps (Figure 2): face acquisition, facial data extraction and representation, and facial expression recognition.

Face acquisition is a processing stage to automatically find the face region for the input images or sequences. Adaboost algorithm is used in this paper, which just detect face in the first frame and then track the face in the remainder of the video sequence. After the face is located, the next step is to extract and represent the facial changes caused by facial expressions. In this paper, the motion features are extracted and described as phase form and then constituted to eigen-sequences, which is useful to the dimension of motion feature. Facial expression recognition is the last stage of facial expression recognition experimental system. The facial changes can be identified as facial action units or prototypic emotional expressions. In this paper, we assume the facial expression sequences satisfy the Gaussian of Mixture model, and use Hidden Markov Model to train and test them.

3. Feature Extraction based on dynamic image sequences

Feature extraction methods can be categorized according to whether they focus on motion and deformation of faces and facial features. Motion extraction approaches directly focus on facial changes occurring due to facial expressions, whereas deformation-based methods do have to rely on neutral face images in order to extract facial features. This paper researches facial expression based on dynamic image sequences, which uses Motion extraction approaches to extract facial motion features between facial expression changing.

The use of optical flow to track motion is advantageous because facial features and skin naturally have a great deal of texture. Using feature vector construction, a low-dimensional weight vector in eigenspace can be obtained to represent the high-dimensional dense flows of each frame. Based on the displacement and weight vectors, the motion information is converted to symbol sequences from which we can recognize facial expressions.

3.1 Motion Feature Extraction

Image processing is performed in two steps. In the first step, a velocity vector is obtained from every two successive frames by using a gradient based optical flow algorithm(B. Horn and B. Schunck, 1981)(Freund Y, Schapire R E., 1997)(Gao Wen, Chen Xilin, 1999). To improve performance, the region for processing is confined to two small regions, one is eye-brow region, and another is mouth region. The angry expression's velocity vector of eye-brow region is as Figure 3. The eye-brow and mouth regions are defined as the characteristic parts in the extraction of features from different facial expressions. These regions were selected based on the results from three-dimensional measurements of expressions, and as such they proved to be the regions in which the changes were most pronounced.

3.2 Feature Vector Construction

In the second step, a feature vector construction processing is applied to a vertical and a horizontal component of the velocity vector field at the regions around an eye and the mouth.

(1)Compute the phase form of a vertical v_k and a horizontal u_k component of the velocity vector field, such as

$$V_k = a_k e^{j\theta_k}, 0 \leq \theta_k < 2\pi.$$

(2) Let R denote the facial area, R_1 denote the eye-brow sub-area, and R_2 denote the mouth sub-area. Let $A_i(R_j)$ denote the i -th group of motion vectors in the sub-area R_j , ($1 \leq j \leq 2$). Grouping of the motion vector is performed according to their orientation:

$$A_i(R_j) = \{V_k | V_k \in R_j, 2\pi i / 8 \leq \theta_k < 2\pi(i+1) / 8\} \tag{1}$$

$$0 \leq i < 8, 1 \leq j \leq 2$$

(3) The energy $E_i(R_j)$, within each direction in sub-area R_j is computed according to the equation:

$$E_i(R_j) = \sum_{V_k \in A_i(R_j)} \|V_k\| \tag{2}$$

The feature vector describing the motion of two consecutive frames is formulated from the set of E_i over the whole area R. The above formulation leads to a 8-dimensional feature vector of averaged powers, within each sub-area and all directions: $E(R_j) = [E_1(R_j), E_2(R_j), \dots, E_8(R_j)]$, $1 \leq j \leq 2$.

4. Facial Expression Recognition based on dynamic image sequences

Facial expression Recognition can be regarded as a pattern recognition problem. It is necessary to model dynamic facial feature vector sequence in order to analyze facial expression sequence. Modeling facial expression needs to take into account the stochastic nature of human facial expression involving both the human mental state, which is hidden or immeasurable, and the human action, which is observable or measurable. For example, different people with the same emotion may exhibit very different facial actions, expression intensities and durations. Individual variations notwithstanding, a human observer can still recognize what emotion is being expressed, indicating that some common element underlies each motion. Therefore, the purpose of facial expression modeling is to uncover the hidden patterns associated with specific expressions from the measured (observable) data. Facial expression modeling requires a criterion for measuring a specific expression. It is desirable to analyze a sequence of images to capture the dynamics. Expressions are recognized in the context of an entire image sequence of arbitrary length. We will develop a recognition system based on the stochastic modeling of the encoded time series describing facial expressions, which should perform well in the spatio-temporal domain, analogous to the human performance.

Some other advantages of using HMMs are: HMM computations converge quickly making it practical for real time processing, it may evaluate an input sequence of uncertain category to present a low output probability, and a multi-dimensional HMM may be developed to integrate individual HMMs to give a robust and reliable recognition.

We usually divide HMM into two types, one is discrete HMM, and another is continuous HMM. As facial expression eigen-sequences are continuous signals. If we use discrete HMM to model eigen-sequences, it might be serious degradation association with vector quantization. Hence it would be advantageous to be able to use HMMs with continuous observation densities.

Continuous HMM model that was based on a probability density distribution of the vector, such models are good or bad depends on the assumption that the probability distribution is consistent with the actual situation. Generally speaking, the probability distribution of some commonly used as Gauss distribution does not accurately describe its distribution, and therefore the Gauss mixed HMM model (Gaussian of Mixture HMM, in short MHMM), which uses several different centres, separated for different combinations of Gauss distribution vector is approaching the actual characteristics of the distribution. Generally speaking, the probability distribution of some commonly used models such as the Gauss distribution MHMM assumptions short periods expression vector characteristics mixed with Gauss distribution. Assuming that the Observer series $O = (O_1, O_2, \dots, O_T)$ expression vector characteristics, whose probability distribution B meet Gauss mixed distribution, the form is:

$$b_j(O_t) = \sum_{m=1}^M c_{jm} \Gamma(O_t, \mu_{jm}, U_{jm}), 1 \leq j \leq N \tag{3}$$

$$\Gamma(O, \mu, U) = \frac{1}{(2\pi)^{\frac{|O|}{2}} \sqrt{|U|}} \exp\left\{-\frac{1}{2}(O-\mu)^T U^{-1}(O-\mu)\right\} \tag{4}$$

where O is the vector being modeled, c_{jm} is the mixture coefficient for the m th mixture in state j and Γ is any log-concave or elliptically symmetric density, with mean vector μ_{jm} and covariance matrix U_{jm} for the m th mixture component in the state j . Usually a Gaussian density is used for Γ . The mixture gains c_{jm} satisfy the stochastic constraint

$$\sum_{m=1}^M c_{jm} = 1 \text{ and } c_{jm} \geq 0; \tag{5}$$

so that the pdf is properly normalized. i.e.,

$$\int_{-\infty}^{+\infty} b_j(x) dx = 1, \quad 1 \leq j \leq N \quad (6)$$

The facial expression image sequence is the same as human's pronunciation sequence characteristic, which has the time order also the irreversible process. So this article uses the left-right MHMM model (1st-order left-right MHMM) to describe the facial sequences. The model structure is as Figure 4. Along with the time increase, each condition only can shift from this condition to the right side neighboring next condition or the maintenance is same, in corresponding transition matrix A only then host diagonal element a_{ii} and the right vice-diagonal element $a_{i,i+1}$ permission non-vanishing, other elements all are the zero.

5. Experiment and Result

We adopt the six universal expressions defined by psychologists Ekman to do experiments, such as angry, disgust, fear, happiness, sadness and surprise. We collected separately for each group of 15 video series, a total of 90 video series. psychology research indicates that typical changes of muscular activities are brief, lasting for a few seconds, but rarely more than five seconds or less than 250ms. So the duration of experiment video series is approximately 3 seconds. The change of expression for the experimental model is neural->apex->neural.

Firstly, we get some special facial expression regions according to the human face structure characteristic. And then carries on the normalization and the standardization. Secondly the motion features are extracted and described as phase form and then constituted to eigen-sequences with the method of part 3.2 described in this paper. Thirdly the eigen-sequences is divided into the training data and the test data, which is used to train the MHMM model parameter and test experiment result separately. We have gathered 90 groups of images sequences. After feature vector construction processing, we get 90 corresponding eigen-sequences. And 60 subassemblies were the training data sets, the other 30 subassemblies for test data set.

To estimate the parameters of a HMM the Baum_ Welch algorithm is used; here, re-estimation is repeatedly performed so as to maximize the generation probability of the training data. For the discrete case a reasonable probability can be obtained from initial values that are set at random. However, for the continuous case the constraints of the parameterized density would limit the range of parameters reachable from a initial value. Therefore, the random setting of initial values will converge into a local maximum in the generation probability, which is far away from the real maximum to be aimed at. To solve this problem, we apply a clustering operation on the training data and the statistical parameters of each cluster are used as an initial value for the parameter of the output probability density. The processing of training HMM parameter is as Figure 5.

In the system, the mix number M of MHMM is defined by the experiment. In order to select optimum mix number M, we have carried on many experiments. The experimental results are as Figure 6. The horizontal axis is mixture number M and the vertical axis is error recognition rate. When mix number M=4, 6, 8, 12, the error recognition rate all might achieve the minimum value. In order to reduce computation complexness, we selected M=4. Therefore, the mixture number M is 4 in this system.

To carry on the expression classification experiment with MHMM, we have designed a MHMM for each kind of expression, altogether six MHMMs, respectively is: anger(1), disgust(2), fear(3), happiness(4), sadness(5) and surprise(6). Each MHMM is a left-right Gaussian of mixture Hidden Markov Model, six such MHMMs constitutes a person face expression Maximum Likelihood classifier, see Figure 2. Given an observation sequence $O = (O_1, O_2, \dots, O_T)$, the probability of the observation given each of the six models $P(O|\lambda_j)$, where $j \in [1,6]$ is computed using the Forward-backward algorithm. The sequence is classified as the emotion corresponding to the model that yielded the highest probability, i.e., $c^* = \text{argmax}[P(O|\lambda_j)]$. We have carried on the test to 30 groups of characteristics sequences, the experimental result is as Table 1.

The experiment proved that, the system in the recognition dynamic expression sequence process, the recognition speed and the recognition result both is ideal. The happiness and disgust expression recognition rate is good, the anger expression recognition rate is relative bad, the overall recognition rate achieves 86.7%.

Use the [6], calculates this sequence separately in six kind of MHMM under model probability $P(O|\lambda_j)$, $j \in [1,6]$. Takes probability value $P(O|\lambda_j)$ the biggest MHMM correspondence expression for this observation sequence expression, namely takes $c^* = \text{argmax}[P(O|\lambda_j)]$ the human face expression serial number which distinguishes for the system.

6. Conclusion

An improved dynamic facial expression recognition experiment system based on the MHMMs model is realized in this paper. And six universal facial expressions: angry, disgust, fear, happiness, sadness and surprise is recognized. We describe facial motion feature based on phase form and construct feature vector, which compress the motion feature and simplified the computing process. In view of the dynamic facial expression analysis, we use Gaussian of mixture model to describe facial expression succession characteristic sequences, which avoided the vector quantification error in the

discrete HMM model. In order to obtain optimum initial value of Gaussian of mixture model, we use k-means algorithm. Which is more precisely describe expression eigen-sequences' probability distribution compares with the traditional model.

But this experimental system is only based on the small scope data acquisition work, uses in training and the test facial expression video frequency sequence is limited. In order to strengthen the experimental system vigorous and healthy, the next step of work may consider the expanded expression database, for example uses the Cohn-Kanade database(Takeo Kanade, Yingli Tian, Jeffrey F. Cohn. 2000). In this experimental system, in view of the different face characteristic region, used the same characteristic number to describe its motion feature, the next step of work might consider using different characteristic number in different face characteristic region separately.

References

B. Fasel, J. Luetttin. (2003). Automatic Facial Expression Analysis: A Survey. *Pattern Recognition*, 2003, 36(1): 259~275.

B. Horn and B. Schunck. (1981). Determining Optical Flow. *Artificial Intelligence*, 1981, 17: 185~203.

Ekman P, Friesen WV. (1978). Facial Action Coding System. Palo Alto: Consulting Psychologists Press Inc., 1978.

Freund Y, Schapire R E. (1997). A Decision-Theoretic Generalization of on-line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 1997, 55(1): 119-139.

Gao Wen, Chen Xilin. (1999). Computer Vision: *Algorithm and System Principle*. Beijing: Tsinghua University Press, 1999. 83~91

J. Lien. (1998). Automatic Recognition of Facial Expression Using Hidden Markov Models and Estimation of Expression Intensity, The Robotics Institute, CMU.

K.Mase. (1991). *Recognition of Facial Expressions From Optical Flow*. IEICE Transactions, 1991, 74(10): 3474~3483.

Rabiner L R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 1989, 77(2): 257- 286.

Takeo Kanade, Yingli Tian, Jeffrey F. Cohn. (2000). *Comprehensive Database for Facial Expression Analysis*. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000: 46~53.

Table 1. The recognition result with MHMMs

	An	Di	Fe	Ha	Sa	Su
An	3	0	0	0	0	0
Di	0	5	0	0	0	0
Fe	0	0	4	0	0	0
Ha	1	0	1	5	0	0
Sa	0	0	0	0	5	1
Su	1	0	0	0	0	4

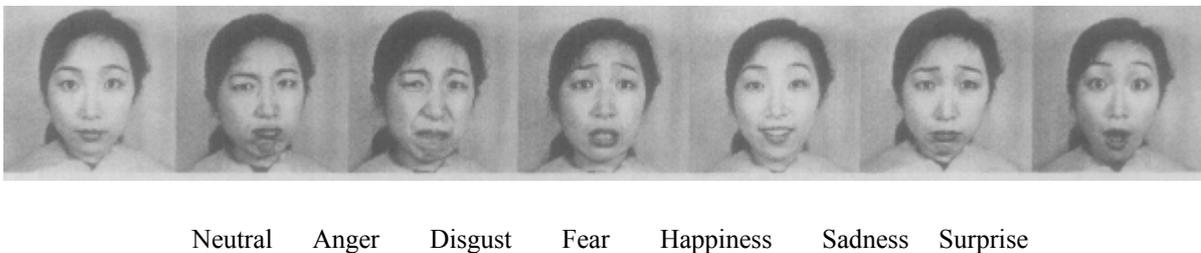


Figure 1. Universal Facial Expressions from JAFFE

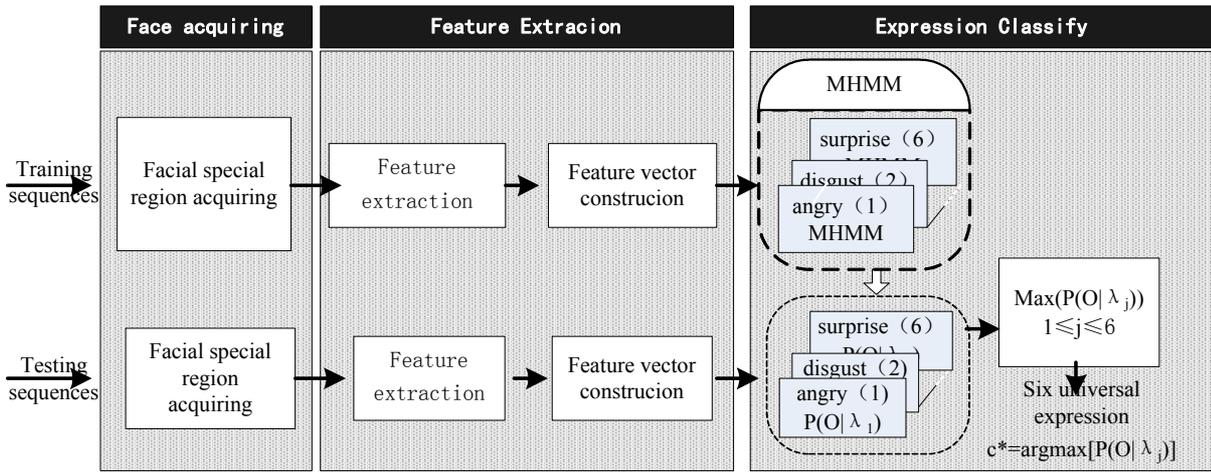


Figure 2. The Basic Structure of Facial Expression Recognition Experimental System

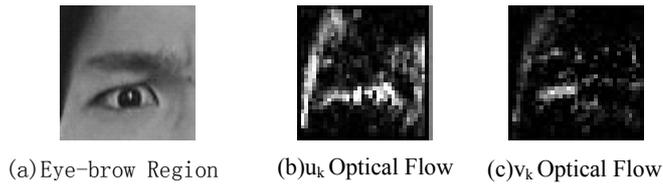


Figure 3. Example of velocity vector

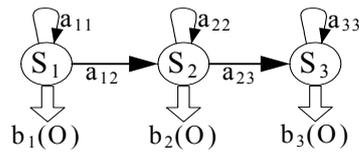


Figure 4. The structure of 1st-order left-right MHMM

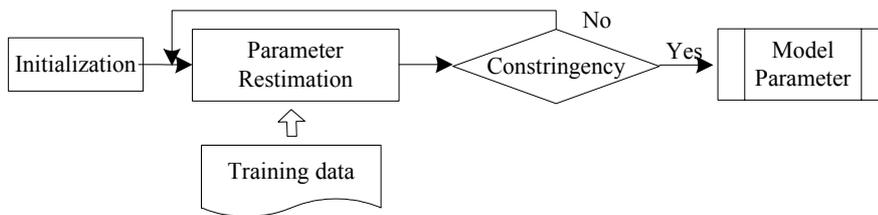


Figure 5. Training MHMM parameter

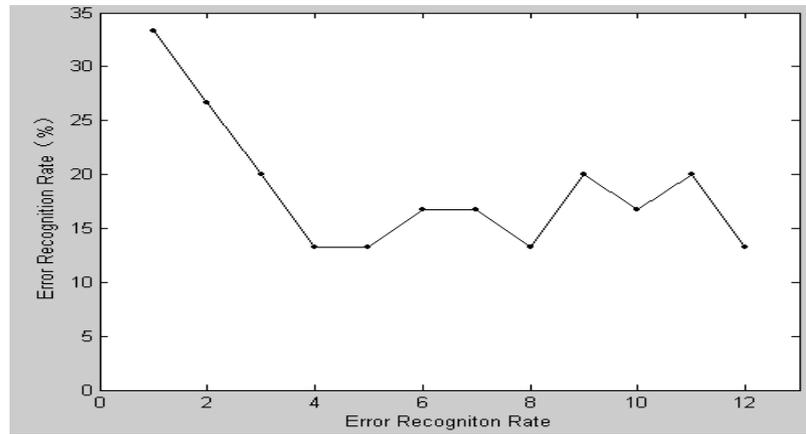


Figure 6. The error rate versus the mixture number M in the MHMMs