Using moment invariants and HMM in facial expression recognition

Y. Zhu, L.C. De Silva *, C.C.Ko

Department of Electrical Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore

Received 16 June 2000; received in revised form 31 May 2001

Abstract

Moment invariants are invariant under shifting, scaling and rotation. They are widely used in pattern recognition because of their discrimination power and robustness. HMM method is a natural and highly reliable way of recognition. In this paper, we have proposed a method of using moment invariants as features and HMM as recognition method in facial expression recognition. Sequences of four universal expressions, i.e. anger, disgust, happiness and surprise, are recognized. We were able to attain an accuracy as high as 96.77%. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Moment invariant; HMM; Pattern recognition; Facial expression; Recognition

1. Introduction

Facial expression recognition is a fundamental technology of advanced communication coding and dynamic human–machine interface systems. The existing facial expression recognition methods are divided into two main categories: static and non-static. Facial expressions are caused by temporal actions of facial muscles. It is usually not robust enough to identify facial expressions just according to static images. This means that temporal information from time sequential facial images should also be exploited.

Facial expressions are reflected by the deformation and displacement of facial features and facial skin. Facial expression recognition can be considered from three perspectives, i.e. characterizing deformation and displacement of facial features, those of facial skin or the combination of them. Mase (1991) used optical flow to estimate the facial skin movements and k-nearest-neighbour as the classification method. Yacoob and Davis (1994) developed a system based on Mase’s work. They used optical flow to estimate facial feature movements and recognized facial expressions according to the relations between facial feature movements and facial expressions. Mase attained an accuracy of nearly 80% when recognizing four expressions: anger, disgust, happiness and surprise. Yacoob and Davis achieved an accuracy of more than 90% to recognize the six universal facial expressions: anger, disgust, fear, happiness, sadness and surprise.
Moment invariants are invariant under scaling, shifting and rotation. They are widely used in pattern recognition. There are several methods of deriving moment invariants. Hu (1962) proposed a method of deriving moment invariants from algebraic invariants and listed examples of different types of moment invariants. Abu-Mostafa and Psaitis (1984) investigated the recognition properties of moment invariants. They studied the influence factors and robustness of moment invariants.

HMM originally appeared in mathematical journals. Rabiner (1989) and Rabiner and Juang (1993) completed the theories of speech recognition using HMM. Sakaguchi et al. (1995) first experienced the effectiveness of employing HMM in facial expression recognition according to image sequences. The features they used are the average powers from distinct frequency bands obtained by applying Wavelet transformation. Discrete HMM is employed as recognition method. They reported 87% recognition rate in user-independent mode using 46 image sequences. Takahiro Otsuka and Jun Ohya (1996) developed a method by using a new feature vector obtained from Wavelet transformation coefficients and continuous HMM was used as the recognition method. Their accuracy was 98% in user-trained mode and 84% in user-independent mode.

In our method we have used moment invariants as features. LBG (Linde et al., 1980) algorithm is employed for vector quantization (VQ) and discrete HMM is used for recognition. Four universal expressions, i.e. anger, disgust, happiness and surprise, are recognized. Some examples are shown in Fig. 1. We have tested the recognition effectiveness of original moment invariants. The recognition rate was 83.87%. We further derived a set of new features by modifying the original moment invariants. It has a better distinguishing ability compared to the original. A recognition rate of 96.77% was achieved with this proposed feature vector.

In Section 2 we introduce our recognition system in detail. In this section, we explain the whole structure of the system, feature extraction method and recognition method. The experiment results and discussion are shown in Section 3. The conclusion we draw is given in Section 4 and the future work is introduced in Section 5.

2. Overview of the facial expression recognition system

2.1. Block diagram of the system

Our facial expression recognition system includes two main modules, i.e. the training module and the recognition module. The feature vectors are transformed to symbol sequences by VQ before entering the training module and recognition module. In the training module, parameters of HMMs corresponding to different expressions are estimated according to the input symbol sequences. When a sequence is recognized, the occurrence probabilities of the sequence under different HMMs are calculated. The HMM having the largest probability corresponds to the recognition result. The block diagram of the proposed recognition system is shown in Fig. 2.

2.2. Feature extraction

2.2.1. Feature vector definition

In this section first we give an overview of the moments and then describe our modified moment invariants.
2.2.2. Review

Ordinary moments: Two-dimensional \((p + q)\)th order moment of a density distribution function \(f(x, y)\) is defined as follows:

\[
m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) \, dx \, dy.
\]  

(1)

Central moments: Two-dimensional \((p + q)\)th order central moment of a density distribution function \(f(x, y)\) is defined as follows:

\[
\mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) \, dx \, dy,
\]  

(2)

where

\[
\bar{x} = m_{10}/m_{00}, \quad \bar{y} = m_{01}/m_{00}.
\]

Central moments and their functions are invariant under shifting.

Moment invariants: Here the details of deriving moment invariants will not be discussed. We present the moment invariants we have used in our facial expression recognition system. They are the general linear transformation moment invariants proposed by Hu (1962):

\[
I_1 = \frac{(\mu_{20}\mu_{02} - \mu_{11}^2)}{\mu_{00}^3},
\]  

(3)

\[
I_2 = \left[ (\mu_{30}\mu_{03} - \mu_{21}\mu_{12})^2 ight. \\
\left. - 4(\mu_{30}\mu_{12} - \mu_{21})(\mu_{21}\mu_{03} - \mu_{12}^2) \right]/\mu_{00}^5,
\]  

(4)

\[
I_3 = \left[ \mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) ight. \\
\left. + \mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2) \right]/\mu_{00}^5,
\]  

(5)

\[
I_4 = \mu_{30}^2\mu_{02}^3 - 6\mu_{30}\mu_{21}\mu_{11}\mu_{02}^2 \\
+ 6\mu_{30}\mu_{12}\mu_{02}^2(2\mu_{11}^2 - \mu_{20}\mu_{02}) \\
+ \mu_{30}\mu_{30}(6\mu_{20}\mu_{11}\mu_{02} - 8\mu_{02}^3) + 9\mu_{21}\mu_{20}\mu_{02}^2 \\
- 18\mu_{21}\mu_{12}\mu_{20}\mu_{11}\mu_{02} \\
+ 6\mu_{21}\mu_{03}\mu_{20}(2\mu_{11}^2 - \mu_{20}\mu_{02}) \\
+ 9\mu_{12}^2\mu_{20}\mu_{02} - 6\mu_{12}\mu_{03}\mu_{11}\mu_{20}^2 + \mu_{03}\mu_{20}^3)/\mu_{00}^7.
\]  

(6)

2.2.3. Modified moment invariants

In pattern recognition, e.g. handwriting recognition, it is desired that the feature vectors do not change under scaling, shifting and rotation. Facial expressions are caused by the actions of facial muscles. The actions of facial muscles change not only the shape of facial features but also their relative positions. So in facial expression recognition displacement information of facial features, such as the relative position of brow and eye and left and right mouth corner, benefits differentiating between different facial expressions. In order to utilize the relative position change between facial features we replace central moments in the formula of moment invariants with ordinary moments, thereby modifying the feature vectors. The elements of the modified feature vector are as follows:

\[
J_1 = (m_{20}m_{02} - m_{11}^2)/\mu_{00}^3,
\]  

(7)

\[
J_2 = \left[ (m_{30}m_{03} - m_{21}m_{12})^2 ight. \\
\left. - 4(m_{30}m_{12} - m_{21})(m_{21}m_{03} - m_{12}^2) \right]/\mu_{00}^5,
\]  

(8)

\[
J_3 = \left[ m_{20}(m_{21}m_{03} - m_{12}^2) - m_{11}(m_{30}m_{03} - m_{21}m_{12}) ight. \\
\left. + m_{02}(m_{30}m_{12} - m_{21}^2) \right]/\mu_{00}^5,
\]  

(9)
\[ J_4 = \left[ m_{30}^2 m_{02}^3 - 6m_{30} m_{21} m_{11} m_{02}^2 \right. \\
+ 6m_{30} m_{12} m_{02} \left( 2m_{11}^2 - m_{20} m_{02} \right) \\
+ m_{30} m_{03} \left( 6m_{20} m_{11} m_{02} - 8m_{02}^3 \right) \\
+ 9m_{21}^2 m_{20} m_{02}^2 - 18m_{21} m_{12} m_{20} m_{11} m_{02} \\
+ 6m_{21} m_{03} m_{20} \left( 2m_{11}^2 - m_{20} m_{02} \right) + 9m_{12}^2 m_{20}^2 m_{02} \\
- 6m_{12} m_{03} m_{11} m_{20}^2 + \frac{m_{03}^2 m_{20}^2}{\mu_{00}} \right] \tag{10} \]

2.3. Feature extraction

We define seven areas on the face for feature extraction. They are left brow, left eye, right brow, right eye, upper mouth, lower mouth and the area between two eyes. Each area we define is of rectangular shape. We denote left brow, left eye, right brow, right eye, upper mouth, lower mouth and the area between two eyes as areas 1, 2, 3, 4, 5, 6 and 7, respectively. The final sizes and centres of the rectangles are adjusted so that each rectangular shape can exactly enclose the whole facial feature in every frame of a sequence. The definition of feature extraction areas is shown in Fig. 3. The feature vectors are as follows:

\[ V_1 = [J_{1,i}], \quad 1 \leqslant i \leqslant 7, \quad j = 1, 2, 3, 4, \]
\[ V_2 = [J_{i,j}], \quad 1 \leqslant i \leqslant 7, \quad j = 1, 2, 3, 4, \tag{11} \]

where \( i \) is the number of the areas, and \( j \) is the number of moment invariants.

When calculating moment invariants, selection of coordinate origins does not influence the results.

While calculating the proposed modified feature vector the result varies with the change of origins. In the case of the sequences we have that the relative positions of two pupils keep in the same place on the face if the subject does not move them deliberately. Therefore we choose the centres of pupils as reference. We define them as the origins for calculating ordinary moments of brow areas and eye areas. This can reflect not only the deformation of brows and eyes but also the relative position change of brows and eyes. In order to represent the deformation of upper mouth and lower mouth most efficiently, the middle of the line connecting left mouth corner and right mouth corner is selected as the origin for calculating ordinary moments of the upper mouth area and the lower mouth area. For nose area, according to the experimental results, we choose the middle of lower bound of the area as origin.

Since the differences of the ranges of \( I_1, I_2, I_3, I_4 \) and \( J_1, J_2, J_3, J_4 \) are large, we use a scaling factor \( \alpha \) to balance the values of \( I_1, I_2, I_3, I_4 \) and \( J_1, J_2, J_3, J_4 \). The definition of \( \alpha \) is,

\[ \alpha = [\alpha_i], \tag{12} \]

\( 1 \leqslant i \leqslant \) the number of elements of the feature vector.

Every element of the feature vector is divided by the corresponding element of \( \alpha \). We also improve the VQ and recognition results by adjusting \( \alpha \). Currently the value of \( \alpha \) is adjusted manually as follows.

First assign an initial value of \( \alpha \) to make all the elements of the feature vector be of order 1. Then adjust \( \alpha_i \)'s corresponding to each feature extraction area separately according to the recognition results. From the recognition results we know the expression which has a low recognition rate and with which expression it is confused. The \( \alpha_i \)'s of the facial areas which change most apparently in the expression and distinguish the confused expressions most are decreased. For example the most active facial area of happiness is mouth. When recognition rate of happiness is low and it is confused with surprise the \( \alpha_i \)'s corresponding to mouth, eyebrows and eyes are decreased.

![Fig. 3. Definition of feature extraction areas: the digits are the numbers of the areas.](image-url)
The feature vector of each frame subtracts that of the first frame which is of neutral state. The resultant feature vectors reflect the deformation and displacement of facial features. They are used in the recognition system. The samples of a happy sequence and a disgust sequence are shown in Fig. 4. The plots of their feature vectors after scaling are shown in Fig. 5. From the plots we can approximately separate the five stages of an expression, i.e. neutral-front, transient-front, peak, transient-back and neutral-back.

2.4. Training and recognition techniques

We use discrete left–right HMM for recognition. Theoretically the 5-state HMM matches 5 stages of an expression best. According to the experimental results the HMM having 3 states and 32 symbols per state has the best recognition accuracy. This is because of the small size of our image database. Thus state 1 of the HMM corresponds to neutral-front stage and about 50% of transient-front stage, state 2 of the HMM corresponds to about 50% of transient-front stage, peak stage and about 50% of transient-back stage, and state 3 of the HMM corresponds to about 50% of transient-back stage and neutral-back stage. The structure of the HMM is shown in Fig. 6.

LBG algorithm (Linde et al., 1980) is employed for VQ. The techniques of applying HMM introduced by Rabiner (1989) and Rabiner and Juang (1993) are used for training HMM and recognition. Baum–Welch method is used for parameter estimation. Viterbi algorithm is used for finding the optimal state sequence. And Forward Procedure is used for calculating the probability of the observed sequence.

3. Experiments and results

The experiments are done under the following conditions:
1. Only the frontal view of the facial image sequences are analysed throughout the whole sequence.
2. The head motion between two consecutive frames is considered small.
3. The subjects are not speaking during image capturing.
4. The subjects do not have facial hair and they are not wearing glasses.

Our image database has 31 image sequences taken from 10 subjects. They are 5 of anger, 7 of disgust, 9 of happiness and 10 of surprise. We use 2 anger, 4 disgust, 4 happiness and 5 surprise to train the HMM.

Our experiments include three parts:
1. Using moment invariants as feature vector.
2. Using modified moment invariants as feature vector.
3. Combining moment invariants and modified moment invariants as feature vector.

The recognition results of using moment invariants and modified moment invariants are shown in Tables 1 and 2, respectively. From the results, we can see that using original moment invariants the results are less accurate. While using the modified moment invariants a recognition accuracy of 96.77% is obtained.

In order to reduce the vector dimension we have also tested by using $J_1$, $J_2$, $J_4$ or $J_1$, $J_3$, $J_4$ as elements of our feature vector. The same recognition rate as that of using $J_1$, $J_2$, $J_3$, $J_4$ was attained.

Fig. 4. Samples of a happy sequence and a disgust sequence (the upper two rows are of a happy sequence and the lower two rows are of a disgust sequence).
Fig. 5. Plots of modified feature vector elements of a happy sequence and a disgust sequence $(I_1, (+); I_2, (+); I_3, (×); I_4, (−))$: (a) is of a happy sequence and (b) is of a disgust sequence.
Fig. 5. (Continued).
Fig. 6. Structure of the HMM used for facial expression recognition results using modified moment invariants $J_1, J_2, J_3$ and $J_6$ or $J_1, J_2, J_3$ or $J_1, J_3, J_4$ or $l_1, l_4$. $J_1 - l_1, J_4 - l_4$.

For further reducing the dimension of feature vector, we have tested using just two of the original four elements. The best combination is $J_1, J_4$. The recognition results are shown in Table 3.

Moment invariants can take the deformation of facial features into account, while modified moment invariants can take both the deformation and displacement of facial features into account. We have tried a simple method to separate the deformations and displacements of facial features, i.e. subtracting original moment invariants from modified moment invariants. The elements of the feature vector are $I_1, I_4, J_1 - I_1, J_4 - I_4$. We were able to obtain the same results as those shown in Table 2.

4. Conclusion

In this paper we have proposed a method of using moment invariants in facial expression recognition. First we directly used a set of general linear transformation moment invariants. Then we derived a modified feature vector from moment invariants. Moment invariants can reflect the deformation of facial features but cannot provide sufficient information of the displacement of facial features, while the modified feature vector reflects deformations and relative displacements of facial features. From the recognition results we can see that directly using moment invariants in facial expression recognition cannot get very satisfactory recognition accuracy. Expressions of anger and disgust cannot be distinguished well. Using our improved modified feature vector a recognition rate as high as 96.77% is obtained. This demonstrates that only using deformation information of facial features is not enough for facial expression recognition. If information of facial feature displacements is also employed, the recognition

| Anger | Disgust | Happiness | Surprise |%
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Disgust</td>
<td>0</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Happiness</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Surprise</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2

| Anger | Disgust | Happiness | Surprise |%
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Disgust</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Happiness</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Surprise</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3

| Anger | Disgust | Happiness | Surprise |%
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Disgust</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Happiness</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Surprise</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
ability can be improved greatly. To reduce the
dimension of feature vector, we have experimented
just using three and two of the four modified
moment invariants. We have found that using $J_1$,
$J_2$, $J_4$ or $J_1$, $J_3$, $J_4$ recognition rate as high as that of
using all the four can be obtained. But just using
two will weaken the recognition ability. We have
also tested using a simple method to separate the
information of displacement and deformation.

5. Future directions

Since the feature extraction areas are selected
manually the recognition results may be influenced
by subjective factors and the bounds of the areas
are invariant during the whole sequence. De Silva
et al. (1995) developed a very efficient way of facial
feature area extracting and feature tracking. In the
future we will add automatic feature tracking
techniques like the one in (De Silva et al., 1995) to
the proposed system for finding the feature ex-
traction areas. Thus the areas can be selected au-
tomatically and more accurately. The recognition
system can be automatic and the recognition re-
sults will be more stable. We also expect to obtain
the analytical conclusions for deciding the values
of $z$ to get the highest accuracy.

References

698–706.

and tracking of facial features by using edge pixel counting
and deformable circular template matching. IEICE Trans.
E78-D (9), 1195–1207.


quantization design. IEEE Trans. Communication Com-28
(1).

Mase, K., 1991. Recognition of facial expression from optical
flow. IEICE Trans. 74, 3474–3483.

Rabiner, L.R., 1989. A Tutorial on hidden markov modes and
selected applications in speech recognition. Proc. IEEE 77
(2), 257–286.

Rabiner, L.R., Juang, B.H., 1993. Fundamentals of Speech
Recognition. Prentic-Hall, Englewood Cliffs, NJ.

Tatsumi Sakaguchi, Shigeo Morishima, et al., Facial Express-
ion Recognition from Image Sequences Using Hidden
Markov Model. ACCV’95, pp. 423–427.

Takahiro Otsuka, Jun Ohy, 1996. Recognition of facial
expressions using HMM with continuous output probabili-
ties. In: IEEE Internat. Workshop on Robot and Human

Yacoob, Y., Davis, L., Computer spatio-temporal represatation
Conf. Computer Vision and Pattern Recognition’94,
pp. 70–75.