VIDEO SEARCHING SYSTEM BASED ON HUMAN FACE IDENTIFICATION AND FACIAL EXPRESSION RECOGNITION USING MSM AND AAM

TOMOKO OKADA*, TETSUYA TAKIGUCHI† and YASUO ARIKI†

*Graduate School of Engineering
Kobe University
1-1 Rokkodai, Nada, Kobe, 657-8501, Japan
e-mail: pekkle@me.cs.scitec.kobe-u.ac.jp

†Organization of Advanced Science and Technology
Kobe University
1-1 Rokkodai, Nada, Kobe, 657-8501, Japan
e-mail: takigu@kobe-u.ac.jp
ariki@kobe-u.ac.jp

Abstract

These days, due to the technological progress being made in digital video devices, demand has been increasing for effective ways to search specific images in large-volume, accumulated recordings. To solve this problem, we propose an efficient video searching system based on human face images. The user will be able to retrieve desired scenes containing the queried person with the specified facial expression in a video file. In the tested system, an actor of a drama is identified and the actor’s facial expression is recognized using the Mutual Subspace Method (MSM) and Active Appearance Models (AAMs). Experiments using drama have yielded results in which the identification rate was 76.8% and the recognition rate of the facial expressions was 84.8%.

Keywords and phrases: content analysis, person identification, facial expression recognition, the mutual subspace method, active appearance models.

Received December 19, 2009
1. Introduction

Recently, a large amount of multimedia information has accumulated in the home due to the rapid spread of digital devices which deal with video, DVD, Blu-ray Disc, and so on. Therefore, it is difficult to search for favorite data from the masses of multimedia information. Specialists have proposed many studies on visual data to solve the problem. For example, Mo et al. [6] proposed the method of classifying many video data based on image features and Miyahara et al. [5] proposed a tagging system for video content based on the facial expressions that can be used for recommendations.

In this paper, we propose an efficient scene retrieval system for video files based on human face images. Since our proposed system can retrieve the desired shots from the video more quickly, the user can watch the scenes where his favorite actor appears with a particular emotional expression. He also can edit the recorded video based on the human face images recorded in it.

For the construction of this system, the Mutual Subspace Method (MSM) [4] and Active Appearance Models (AAMs) [1] are employed. The MSM is used to identify the person and AAM is used to extract facial feature points for recognizing facial expressions.

The rest of this paper is organized as follows. Section 2 gives the overview of our system. Sections 3 and 4 describe the proposed methods. Section 5 describes experiments and evaluations. Section 6 discusses future themes of study.

2. Overview of the Proposed System

2.1. Methods used for the proposed system

In this paper, a TV drama is targeted for experiments. The actor moves and performs freely in the drama. Since the actor’s face direction and facial expressions are variable, it is necessary to employ robust methods for identifying the actor and recognizing his facial expressions. From this point of view, we employ MSM to identify the actor and AAM to track the actor’s face. Thanks to these methods, we can identify the person and recognize his facial expressions precisely even if he moves freely in the video.

2.2. Flow of the proposed system

Figure 1 shows the flow of our system. At first, exact face regions are extracted
from the input face image sequence in the video by AdaBoost based on Haar-like features [8]. After this, these face images are compared with a face database for identification using MSM. Next, facial feature points are extracted with AAM, and the frontal facial images are generated with a computation of the face direction [2]. Finally, the actor’s facial expression is recognized by the classifier, Support Vector Machines (SVMs) [7].

3. Identifying the Person Using MSM

MSM can compute the similarity between multiple images obtained as training data and multiple input images fed as test data. This method is mainly used for face recognition and character recognition. Since MSM constructs the subspaces from test data as well as training data, it is robust for the change of face direction and the facial expressions. Figure 1 shows the configuration of MSM.

![Figure 1. The flow of the proposed system.](image1)

![Figure 2. The configuration of MSM.](image2)
For the comparison between subspaces $\mathcal{P}$ and $\mathcal{Q}$, the similarity $\theta$ between them is computed in formula (1). Vectors $u$ and $v$ are projected vectors to subspaces $\mathcal{P}$ and $\mathcal{Q}$, respectively,

$$
\cos^2 \theta = \max_{u \in \mathcal{P}, v \in \mathcal{Q}} \frac{\|u\|\|v\|}{\|u \cdot v\|}
$$

(1)

4. Extraction of Facial Feature Points and Generation of Frontal View Face

AAM is used in many applications such as facial parts tracking. It is a statistical model that shows the correlation between the shape (coordinate values of feature points) and the texture (intensity of each pixel). Since dimensions of these features are reduced by PCA, using AAM can provide fast and stable object tracking. The structure and search method of AAM and frontal view face generation using AAM are described below.

4.1. The structure of AAM

Figure 3. The structure of AAM.

Figure 3 shows the structure of AAM. In this paper, a vector composed of coordinate values of the feature points is called $shape vector s$. To construct AAM,
the face region is extracted from images with its feature points. Then, its shape is
normalized into a mean shape \( \bar{s} \). A vector composed of intensity values in the
extracted images is called \textit{texture vector} \( g \). PCA is performed on a set of shape
vector \( s \) and texture vector \( g \) in the training data. Formula (2) gives
\[
s = \bar{s} + \Phi_s b_s
\]
and
\[
g = \bar{g} + \Phi_g b_g \tag{2}
\]
where \( \bar{s} \) is a mean shape vector of \( s \) and \( \bar{g} \) is a mean texture vector of \( g \). \( \Phi_s \) and
\( \Phi_g \) are orthogonal matrices whose columns are basis vectors and \( b_s \) and \( b_g \) are
coefficient vectors to the basis vectors. Then, the shape and the texture coefficient
vectors \( b_s \) and \( b_g \) are concatenated after \( b_s \) is weighted by matrix \( W_s \) and PCA is
performed again as shown in formula (3): Here, \( c \) and \( Q \) are parameter vectors
controlling both the shape and the texture, and the orthogonal matrix, respectively.
After this, units of the shape and the texture are normalized by matrix \( W_s \) and PCA
is performed again. \( c \) is a parameter vector controlling both the shape and the
texture. \( Q \) is an orthogonal matrix which corresponds to vector \( c \). This is given in
formula (3),
\[
b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s \Phi_s^T (s - \bar{s}) \\ \Phi_g^T (g - \bar{g}) \end{pmatrix} = Qc. \tag{3}
\]

4.2. Searching method of AAM

When a new input image is given, AAM constructed in the way as mentioned
above is translated, rotated and synthesized with a parameter vector \( c \). Then, the face
is tracked by a minimization of the texture difference between the new input image
and a synthesized image as follows:
\[
c^* = \arg \min_c \| g_n - g_s \|^2, \tag{4}
\]
where \( g_s \) is the texture of the synthesized image reconstructed from parameter \( c \) and
\( g_n \) is the texture of the new input image. The facial feature points are extracted by
searching the value of the parameter \( c^* \) of AAM, which is the most similar to the
new input image.
4.3. Frontal view face generation using AAM

Frontal view face generation makes it possible to recognize facial expressions in any face direction and to reduce the amount of training data. Two characteristics of the model vector \( c \) obtained by AAM are used. One characteristic is that the low component of \( c \) contains information about the face direction. The other is that the face direction \( \theta \) and the model parameter \( c \) are highly correlated to each other [2]. This is described as follows:

\[
\theta = d0 + d1 \ast \theta,
\]  \hspace{1cm} (5)

where \( d \) is the low component of \( d \), and \( d0 \) and \( d1 \) are constant vectors learned from the training data in accordance with the least mean square method.

Given a face image with a parameter \( d' \), we can estimate the face direction \( \theta' \) as follows:

\[
\theta' = (d' - d0)/d1 \hspace{1cm} (d1 \neq 0).
\]  \hspace{1cm} (6)

To generate frontal view face images, the frontal direction, namely \( \theta = 0 \), is substituted into formula (5),

\[
d_{\text{front}} = d0 + d_{\text{res}}. \hspace{1cm} (7)
\]

Thus, we can generate the frontal view images using the above \( d_{\text{front}} \).

5. Experiments

5.1. Experimental conditions

To evaluate our proposed system, we used a TV drama that was 20 minutes long. The frames where an actor appeared were tagged as the ground truth for the experiment of identifying the actor. The training data for MSM and AAM were collected from a TV drama that was 40 minutes long and different from the tagged drama. The facial expression to be recognized was smile.

5.2. Experimental results

The results of the identification are shown in Table 1 and the results of the facial expression recognition are shown in Table 2. Table 2 shows the results of three types of experiments; \( \alpha \) (successful identification and facial expression recognition for all frames), \( \beta \) (successful facial expression recognition irrespective of identification
success for all frames), \( \gamma \) (successful facial expression recognition for the frames of successful identification). From the results of the experiments, it can be seen that the identified rate is about 76.8% and the rate of the facial expression recognition is about 84.8%. The main reason that the recognition rate is bounded below about 85% can be attributed to the tracking accuracy by AAM due to the variation of conditions in the video such as illumination.

<table>
<thead>
<tr>
<th>Table 1. Results of the identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>0.790</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Results of the facial expression recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of experiments</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>( \alpha )</td>
</tr>
<tr>
<td>( \beta )</td>
</tr>
<tr>
<td>( \gamma )</td>
</tr>
</tbody>
</table>

6. Conclusion

In this paper, we proposed an efficient video searching system based on the human face identification and facial expression recognition. As shown in the experimental results, the identification rate was about 78% and the rate of the facial expression recognition was about 85%. Based on these rates, it can be said that our system works well. In the future, we will improve the accuracy of identification and facial expression recognition, as well as increase the classes of facial expressions.

References


