Face Detection and Recognition using Colour Sequential Images
Zhonglong Zheng
Institute of Information Science and Engineering, Zhejiang Normal University, Jinhua, 321004, Zhejiang, China
E-mail: zhonglong@sjtu.edu.cn; zhonglong_qd@163.com
Jie Yang
Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, 200030, China
Yitan Zhu
Alexandria Research Institute, Virginia Polytechnic Institute And State University, 206 N. Washington Street, Alexandria, VA 22314, USA

A human face detection and recognition system for colour sequential images is presented in this paper. The system is composed of two subsystems: human face detection subsystem, human face recognition subsystem. The face detection subsystem includes two modules: face finding, and face verification. The human face finding module finds the candidate face regions from colour sequential images according to skin colour analysis and motion analysis. The human face verification module has been developed to verify the detected human faces by judging of eclipse and Support Vector Machines (SVM), and precisely localize human faces by locating eyes and mouths based on Generalized Symmetry Transform. The features of the relation between face patterns can be extracted and selected by Principal Component Analysis. The selected features are used to train multiple SVMs which can finally classify human faces. The system structure is designed according to the following principle: Firstly simpler methods are used to reduce the search space, and then more complicated methods are used in the reduced space. So the system can have a quick response speed as well as holding high detection and recognition rate. Human face detection accuracy of the system is 97.2% under controllable lightning condition. Human face (70 persons) recognition accuracy of the system is 96.5% (with 20 eigenvectors) and 98.3% (with 30 eigenvectors).

ACM Classification: I.5.4 Pattern Recognition

1. INTRODUCTION
Face detection and recognition, as a specific content of pattern recognition, has a wide range of applications in virtual reality, intelligent human machine interface, videophone systems and security systems, etc. A successful and applicable system for human face detection and recognition will have a great effect on human society. An integrated human face detection and recognition system consists of many modules, such as face detection/location, feature extraction/selection, face database organization and face recognition. A typical procedure of human detection and recognition is shown in Figure 1.

Copyright© 2006, Australian Computer Society Inc. General permission to republish, but not for profit, all or part of this material is granted, provided that the JRPIT copyright notice is given and that reference is made to the publication, to its date of issue, and to the fact that reprinting privileges were granted by permission of the Australian Computer Society Inc.
Manuscript received: 19 January 2005
Communicating Editor: Geoff McLachlan
Face Detection and Recognition using Colour Sequential Images

This system is based on colour sequential images, because it has a better application value in the future. For a successful face detection and recognition system, face detection is a very important stage. It directly affects other modules, such as feature extraction and face recognition.

In recent years, more attention has been paid to face detection, and related research. The current processing methods include greyscale processing, chroma chart, motion analysis etc. Face recognition modules play an important role in the total system. Traditional methods include, appearance-based algorithms (Turk and Pentland, 1991), geometrical–based algorithm (Brunelli and Poggio, 1992), 3D-based algorithms (Blanz, 2003), neural network methods (Rowley et al, 1998) and so on.

All the methods mentioned above have achieved good performance in particular conditions. But the pattern of the human face is both particular and complicated because of the following factors. Information about the face is of high dimension and is hard to model. The face is a non-rigid object, which contains great variation. The factors such as lighting, emotion, visual angle, age, accouterment, and hair affect the pattern greatly, and make the pattern nonlinear. Up to now a robust method which can recognize the human face in any scene has not been found.

Real applications are limited, as people cannot collect mass samples for each human face. Compared with its high dimension of features, the number of face samples is relatively small, so human face recognition becomes a small sample problem. For small sample problems, traditional methods, such as neural networks, have an inborn shortcoming: poor generalization. The other disadvantage of traditional methods is a learning ability which cannot effectively classify very complicated patterns such as human faces. The Support Vector Machine (SVM) introduced by Vapnik (1998) is based on statistical learning theory. It has following advantages: 1) Traditional minimum empirical risk is replaced by structural risk. 2) It has both good learning ability and generalization ability. 3) It can overcome the phenomenon of overfitting. SVM is very useful in applications solving small sample problems and nonlinear classification. SVM is used in our system for face verification and face recognition.

In this paper, a human face detection and recognition system in colour sequential images is presented. Colour sequential images are taken by a colour camera. The system can detect and track human faces automatically, and can recognize human faces in real-time. The system automatically gives an alarm when an unauthorized person is found. The requirements of the application environment of the system are: controllable in-house lighting, and front view human face image. The system is composed of two subsystems: a human face detection subsystem and a human face recognition subsystem. The human face detection subsystem is described in Section 2. The human face recognition subsystem is described in Section 3. The conclusions and future work are discussed in Section 4.
Face Detection and Recognition using Colour Sequential Images

2. HUMAN FACE DETECTION
The face detection subsystem includes two modules: face finding and face verification. Figure 2 is the flowchart for the human face detection subsystem.

2.1 Human face finding
The human face finding module finds the candidate’s face regions from colour sequential images according to skin colour analysis and motion analysis. The human face finding module is composed of skin detection (Wu and Cai, 1999), motion detection (Crowley *et al*, 1997) and horizontal adjustment.

2.1.1 Skin detection
In most cases, images are stored in a computer in RGB format. But since each component of RGB holds data on lightness, a relation exists in RGB components. So before skin detection, the RGB format should be converted to the UCS (Uniform Chromatic Scale) colour system, which is defined by CIE. The conversion from RGB to UCS is shown as follows:

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} =
\begin{bmatrix}
0.490 & 0.310 & 0.200 \\
0.177 & 0.813 & 0.011 \\
0.000 & 0.010 & 0.990
\end{bmatrix}
\begin{bmatrix}
R_{CIE} \\
G_{CIE} \\
B_{CIE}
\end{bmatrix}
\]

(1)
Face Detection and Recognition using Colour Sequential Images

\[ u = 4X/(X + 15Y + 3Z) \quad \quad v = 6Y/(X + 15Y + 3Z) \]  

(2)

In the UCS system, \( Y \) stands for lightness, \( u \) and \( v \) stand for hue, and \( u \) and \( v \) are independent. Since \( u \in [0,1], v \in [0,1] \), we count a great deal of skin points (about 10,000), then divide \( u \) and \( v \) into 100 pieces, and represent the statistical distribution of \( u \) and \( v \) by the matrix. Considering the following function:

\[ P(x, y) = N((i-1) < x \leq i, (j-1) < y \leq j)/N \]  

(3)

In equation (3), \( N((i-1) < x \leq i, (j-1) < y \leq j) \) stands for the number of skin points in \( i \)th, \( j \)th piece, \( N \) stands for the total number of skin points, and \( 0 \leq i \leq 100, 0 \leq j \leq 100 \). We call \( P(x, y) \) the skin hue \( uv \) statistical surface, which represents the statistical distribution of skin hue in hue space.

We treat the \( uv \) statistical surface as a mapping function. A multi-layer perceptron network is adopted to approximate the skin colour distribution in the UCS colour space. The advantages are that it is adapted to various lightning conditions while remaining robust even in complex background situations, the optimum parameters are adjusted automatically, and a high face detection rate is achieved.

Detection of skin region: According to the skin colour distribution approximated by the Multi-layer perceptron network, each RGB image can be mapped to a 256 greyscale statistical image. A high density point has a high probability of being a skin point, otherwise the probability is low. The greyscale image is segmented by selecting an appropriate threshold \( T_s \) (the default threshold is 128). Then a binary image is obtained, and the skin region can be extracted by a region dilation method. A morphologic filter is used to remove region noise. At least \( k \) skin regions \( r_1, r_2, \ldots, r_k \) can be obtained, which are called the aggregate of skin region \( R_c \). \( R_c = R_{cm} \cap R_{cs} \), \( R_{cm} \) is the motion skin region while \( R_{cs} \) is the static skin region.

2.1.2 Motion detection and analysis

Since motion analysis is based on greyscale images, we firstly convert the RGB image to a 256 level greyscale image, and the difference between two neighbouring images can reflect the motion. The difference is defined as follows:

\[ f_{\text{diff}}(x, t_1, t_2) = f(x, t_1) - f(x, t_2) \]  

(4)

In equation (4), \( f_{\text{diff}} \) is the differential image, \( f(x, t_1) \) is the image at time \( t_1 \) while \( f(x, t_2) \) is the image at time \( t_2 \). The difference image holds the edge information of the original image of moving objects.

We count the average \( m_i \) for each skin region \( R_c \) in image \( f_{\text{diff}} \) :

\[ m_i = \frac{1}{N_i} \sum_{x \in R_i} f_{\text{diff}}(x) \]  

(5)

In equation (5), \( r_i \) stands for the \( i \)th skin region of \( R_c \), \( N_i \) stands for the number of points in region \( r_i \), \( m_i \) stands for average difference value of region \( r_i \).

By selecting the appropriate threshold \( T_m \) (the default value is 8, which depends on motion speed, lightness, and can be adjusted dynamically), when \( m_i \) of region is \( r_i \) less than \( T_m \), it is classified as a static region, otherwise, it is classified as a motion region. By motion detection, \( R_{cs} \) can be removed from \( R_c \).

2.1.3 Horizontal adjustment

As the motion image is derived from the differential image, when human faces in colour sequential images move in a horizontal direction, candidate regions detected by motion analysis have
excursions in a horizontal direction. So the horizontal adjustment module is necessary, which adjusts candidate regions based on the symmetry of the skin probability image. After horizontal adjustment, human faces are basically located in the centre of candidate regions, and few background regions are included in each candidate region.

Figure 3 is the experimental result of the face finding module, the face finding module can scan and find more than one face in colour sequential images.

2.2 Human face verification

On the first frame, as we do not know the position of the human face, an exhaustive search should be carried out. Firstly we analyze the skin hue, and the region that does not match the skin hue can be removed from the image. The remaining region is similar to skin, such as face, arm and other similar background objects. The next step is motion detection to remove static regions. After combined motion analysis with skin analysis, only regions that are both moving and similar to skin remain. The above procedure has removed a large majority of no-face regions, but fake regions may remain, such as a moving hand. We believe that any kind of human face detection algorithm will be inherently unstable if it does not have a face verification component. The human face verification module is developed to verify the detected human faces and precisely localize human faces.

Since the shape of a human face is like an ellipse, first verification is done by judging every $r_i$ in $R_{cm}$ if it is an ellipse, we can remove the non-face region from $R_{cm}$. For example, a moving arm will be removed because of its long shape.

We implement the algorithm as follows: Firstly each $r_i$ is filled to become a non-concave region $r_i'$. If $|S(r_i') - S(r_i)| / r_i > 0.2$, $r_i$ is considered to be a non-ellipse, here $S(r)$ stands for the square...
Face Detection and Recognition using Colour Sequential Images

of region \( r' \). If \( |S(r') - S(r)|/r \leq 0.2 \), then if \( \pi \leq l^2_r/4S(r') \leq 4.25 \), the region is considered to be a human face, otherwise it is considered as a non-face, here \( l_r \) stands for perimeter of region \( r' \).

But there still exists some non-face regions that cannot be removed by judging non-eclipse, two SVMs are used to carry out final verification. The output of a SVM is +1 and −1, +1 stands for face region while −1 stands for non-face region.

2.2.1 Verification based on Support Vector Machine

The procedure of face verification based on SVM is: 1) Selection of face region (Figure 4); 2) Digital Cosine Transformation on a face image, and selection of low frequency coefficients; 3) Division of face region into several sub-regions (Figure 5); 4) Selection of negative samples 5) Training SVM.

The requirements of negative sample selection are: 1) As many as possible; 2) Correlation between samples should be small; 3) Including both the samples that are like face and are not like human face. Images of negative samples should be produced randomly and automatically, and contain rich information including many kinds of patterns. One kind of candidate images used as negative samples is landscape. As the two eye blocks are symmetric, only one 39 dimensional SVM is needed, chin block is relatively simple, it can be described by one 33 dimensional SVM. There are 1400 positive samples (70 persons, each with 20 images) and more than 5000 negative samples which are automatically produced and used for training the two SVMs. Finally the two trained SVMs can be used to verify a human face, if an SVM judges the input image to be a human face by outputting +1, to be a non-face image by outputting −1.

2.2.2 Location of human face based on the geometric features

The aim of facial feature localization is to provide helpful information for face verification, recognition of face modelling (Vezhnevets and Degtiareva, 2003; Taro, 2002; Feng 2001). Because the features of nose are not easily detected in a robust way, we do not locate the nose. With the locations of eyes and mouth, the human face can be more precisely located.

Generalized Symmetry Transform (GST) is a method to describe symmetry of points. Since the centres of eyes always have the highest symmetry in the human face, we can use GST to locate
them. Because this method only uses the biometrics distribution features of the human face, it is more robust than some other methods under the variation of illumination, pose and expression.

**Definition of GST**

GST performs local operations on the edges of the image. It assigns a continuous symmetry measurement to each point in the image, rather than a binary symmetry label (Reisfeld *et al.*, 1996). We first define a symmetry measure for each point. Let $p_k = (x_k, y_k)$ be any point ($k = 1, 2, \ldots, K$), and denote by $\nabla p_k = (\frac{\partial p_k}{\partial x}, \frac{\partial p_k}{\partial y})$ the gradient of the intensity at point $p_k$. We assume that a vector $v_k = (r_k, \theta_k)$ is associated with each $p_k$ so that $r_k = \|\nabla p_k\|$ and $\theta_k = \arctan\left(\frac{\partial p_k}{\partial x}, \frac{\partial p_k}{\partial y}\right)$. For any two points $p_i$ and $p_j$, $l$ denotes the line passing through them, and $\alpha_{ij}$ denotes the angle counterclockwise between $l$ and the horizon. We define the set $\Gamma(p)$, a distance weight function $D_\sigma(i,j)$, and a phase weight function $P(i,j)$ as follows:

$$\Gamma(p) \equiv \left\{ (i,j) \left| \frac{p_i + p_j}{2} = p \right. \right\}$$

$$D_\sigma(i, j) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{|p_i - p_j|^2}{2\sigma^2}}$$

$$P(i, j) = (1 - \cos(\theta_i + \theta_j - 2\alpha_{ij}))(1 - \cos(\theta_i - \theta_j))$$

We define the contribution of the points $p_i$ and $p_j$ (Figure 6) as $C(i,j) = D_\sigma(i,j)P(i,j)r_ir_j$. This measurement can be easily normalized, and reflects the fact that each of its components modulates the others. The symmetry magnitude or isotropic symmetry $M_\sigma(p)$ of each point $p$ is defined as

$$M_\sigma(p) = \sum_{(i,j) \in \Gamma(p)} C(i,j)$$

which averages the symmetry value overall orientations.

**Figure 6: The contribution to symmetry of the gradients at $p_i$ and $p_j$**
Face Detection and Recognition using Colour Sequential Images

Algorithm of Locating Eyes and Lips

GST and its extended methods, such as Direction Symmetry Transform and Discrete Symmetry Transform, have already been used in analyzing face images (Reisfeld et al, 1998). But the existing methods are computationally expensive. The reason for this trouble is that different dimension factors will detect different symmetry centres. In other words, a specific dimension factor can only detect symmetry centres whose dimension is similar to it. So if we use the size of eyes as $\sigma$, we can find the centres of eyes. Because no dimension information is known in these existing methods, they have to compute a large range of dimensions. But in our method, the size of the eyes can be estimated according to the face detection results, so only a small range of dimensions need to be computed. Experimental results show that the suitable range is 1/12~1/10 of the width of face image.

The following three steps can localize eyes in a grey face image. Firstly, we use Canny operator to detect the edges of the grey image (Figure 7 (a)) (Liu, 2002). The processing result is an edge image (Figure 7 (b)). Secondly, the symmetry of every point in the edge image is calculated by GST (Figure 7 (c)). In the process of calculating symmetry of every point, there is a problem similar to the bound effect in convolution. It is how to calculate the symmetry of points where distances to the bound of the face image are smaller than $\sigma$. Since the eyes are seldom located in this boundary region, we can set 0 as the symmetry of these points and do not need to calculate their symmetry. Thirdly, 4~7 points with the highest symmetry are selected out to locate the centres of eyes (see Figure 7 (d)). The exact number of selected points is determined based on experiment experience. If too few points are selected, they may locate in the region of one eye. If too many points are selected, some points with high symmetry, which are not in the regions of eyes, may also be selected.

The following is our algorithm to locate the mouth. Firstly, with GST locating eyes on its grey image, Figure 8 (b) is obtained from Figure 8 (a). Secondly, a line segment is used to connect the centres of two eyes. Then the perpendicular bisector of the line segment and two parallel dashed lines (Figure 8 (c)) are drawn. The distance from each of these two dashed lines to the perpendicular bisector is 1/10 of the width of face image. So the width of the stripe region, which is between the two dashed lines, is 1/5 of the width of the face images. Thirdly, the average grey level of every row in the stripe region on the lip probability image is calculated. Thus the vertical distribution of average grey level is obtained (Figure 8 (d)). Fourthly, the vertical coordinate of the row where the
grey level is the highest is denoted by $y_{\text{max}}$. The point on the perpendicular bisector where the vertical coordinate is $y_{\text{max}}$ is the centre of the mouth (see Figure 8 (e)).

After the location of eyes and mouth, we can extract the face from the face image with the template shown in Figure 9 (e.g. see Maio et al., 2000). Figure 10 shows the detection results by localizing facial features.

Figure 8 Localizing mouth

Figure 9: The template that is used to locate face according to the locations of eyes and mouth
3. HUMAN FACE RECOGNITION

The database used in this paper is the Shanghai Jiaotong University (SJTU) human face database (Figure 11), which contains 70 persons, each with more than 20 images of a different viewpoint and emotion. All images are taken under uniform lightning conditions. All the face images are high resolution and can be used for human face detection, recognition and can meet the requirement of human face emotion analysis. All the images are 24bit true colour JPEG images with a resolution 600 × 800. The image index is described in Table 1.

3.1 Feature extraction and selection by PCA

The human face is a high dimensional visual pattern, a greyscale image of 120 × 160 has 19,200 data points. It is incredibly difficult to directly process on a human face image for recognition and classification.

<table>
<thead>
<tr>
<th>Image index</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viewpoint</td>
<td>Front View 1</td>
<td>Front View 2</td>
<td>Left 5º</td>
<td>Left 10º</td>
<td>Left 15º</td>
</tr>
<tr>
<td>Image index</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Viewpoint</td>
<td>Left 20º</td>
<td>Right 5º</td>
<td>Right10º</td>
<td>Right15º</td>
<td>Right20º</td>
</tr>
<tr>
<td>Image index</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Viewpoint</td>
<td>Upward 5º</td>
<td>Upward 10º</td>
<td>Upward 5º</td>
<td>Upward 10º</td>
<td>Speaking 1</td>
</tr>
<tr>
<td>Image index</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>Emotion</td>
<td>Speaking2</td>
<td>Strict 1</td>
<td>Strict 2</td>
<td>Relax 1</td>
<td>Relax 2</td>
</tr>
</tbody>
</table>

Table 1: Human face database with different viewpoint and emotion
A discrete K-L transform method is used in this paper. The K-L transform method is first introduced for the optimal representation for human face image (see Turk and Pentland, 1991). A human face image can be approximated by only 20–40 parameters. Further research has shown its ability for human face recognition.

Each row of the image is concatenated to form a long vector, and the following computation is performed

\[ \Sigma_X = \frac{1}{M} \sum_{i=1}^{M} (x_i - m)(x_i - m)^T / M \]  

where \( \Sigma_X \) is the covariance matrix, \( X \in \mathbb{R}^N \) is the long vector of each face image, \( m \) is the average vector.

Compute the eigenvectors and eigenvalues of \( \Sigma_X \), and sort in descending order. By only selecting 20–40 eigenvectors (namely eigenfaces, Figure 12) which correspond to maximum eigenvalues, an energy ratio of 70–80% can be reached. Every human face image can be represented by a linear combination of the eigenvectors, which can be represented by weights of the vectors. So the human face can be represented by a 20–40 dimensional weight.

As the discrete K-L transform does not have a fast algorithm for increasing the response speed and scalability of the system, the image database is divided into three sections: female section (22 persons), long face section (22 persons) and oval face section (26 persons). The K-L transform computation is done not only on the total image database, nor on each person, but on each section.

Figure 12: Human faces and their eigenfaces, the first of right is the average image
3.2 SVM Training
Due to practical considerations, we cannot collect a large number of face images for every person; a reasonable number will be below 20 images. Compared with its dimension, the number of samples for each person is low, human face recognition becomes a small sample problem.

SVM is capable of dealing with small sample problems, it can overcome the disadvantage of overfitting of traditional methods such as neural networks, it has also the strong classification ability of nonlinear patterns.

But the general SVM can only classify 2-class problems, and cannot classify multi-classes. Each person has one trained SVM for human face recognition. Assume that the human face database has \( n \) persons, SVM training is implemented as follows:

- All the samples of the first person are labeled as (a), and all the samples of other persons are labeled as (b), then all the labeled samples are treated as input samples to train an SVM. Support vectors and classification plane are computed, then the SVM are labeled as ➀, which can classify the first person from other persons.
- All the samples of the second person are labeled as (\( a' \)), and all the samples of other persons are labeled as (\( b' \)). The next operations are the same as step 1. Finally the trained SVM ➋ can classify the second person from other persons.
- Repeat the above steps for all the persons in the human face database, finally \( n \) SVMs are produced.

3.3 Face recognition by SVM
The trained SVMs can classify \( n \) persons, not only two persons. When a new testing sample is input into the \( n \) SVMs, the output of face recognition will be as follows:

- The \( i \)th SVM classifies the input sample to class 1, and the other SVMs classify it to class 2, then the input sample is recognized as the \( i \)th person.
- If more than one SVM classify the input samples to class 1, and other SVMs classify it to class 2, then an error occurs.
- All the SVMs classify the input sample to class 2, then the input sample is judged as a new person that does not belong to the image database.

3.4 Analysis and discussion of experiment results
SVM has good learning and generalization ability for small sample problems, compared with traditional methods. It shows higher face recognition ability.

Table 2 is the comparison of experiment results of classification of the human face images that are represented by 20 dimension eigenvectors. We have discovered that, compared with traditional

<table>
<thead>
<tr>
<th>Classification method</th>
<th>Accuracy of female section</th>
<th>Accuracy of long face section</th>
<th>Accuracy of oval face section</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.9609</td>
<td>0.9649</td>
<td>0.9745</td>
</tr>
<tr>
<td>K-Mean</td>
<td>0.9348</td>
<td>0.9079</td>
<td>0.9562</td>
</tr>
<tr>
<td>Euclidean distance</td>
<td>0.9261</td>
<td>0.8965</td>
<td>0.9343 *</td>
</tr>
<tr>
<td>Mahalanobis distance</td>
<td>0.9261</td>
<td>0.8965</td>
<td>0.9432</td>
</tr>
<tr>
<td>Neural networks</td>
<td>0.9410</td>
<td>0.9221</td>
<td>0.9218</td>
</tr>
</tbody>
</table>

Table 2: The comparisons among SVM and traditional methods
face detection and recognition using colour sequential images

Methods, such as K-Mean, Euclidean distance, Mahalanobis distance, and neural networks, SVM achieves the highest accuracy. Meanwhile, when the face images are represented by 25 and 30 dimension eigenvectors, SVM has a classification rate of nearly 100\%, while the traditional method is 96\%.

The performances of different kernel functions of SVMs are also compared by experiments (Table 3). It is concluded that 2-order polynomial SVM have a better performance when the face image is represented by few eigenvectors, with higher recognition rate and fewer support vectors.

### 4. CONCLUSIONS AND FURTHER RESEARCH

Skin chroma combined with motion analysis is proposed to quickly detect the human face in colour sequential images, and the detected face regions are verified by SVM and the geometric features. A high face detection rate has been achieved. The combination of these two components can detect faces in colour sequential images efficiently. Advantages: The two components are relatively independent; each can be improved without affecting the other one. A robust PCA method is used to extract features, and SVMs are used to recognize the human face, the face recognition rate is higher than traditional methods. There are two advantages to this system: One is that the system is divided into many steps, the former steps have less computation than later steps, and the former steps narrow the research space for the later steps, which only need to compute on a subset, then the total computation is decreased, system response speed is greatly increased. The second is the focus on the features (complex pattern but few samples) of the human face, the SVM method is used to recognize the human face, then the inherent disadvantages (overfitting) of traditional methods have been overcome, and a higher recognition rate is achieved.

The performance of our system is as follows:

<table>
<thead>
<tr>
<th>Number of eigenvectors</th>
<th>Kernel functions of SVM</th>
<th>Accuracy of female section</th>
<th>Accuracy of long face section</th>
<th>Accuracy of oval face section</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>RBF</td>
<td>0.9826/39.3</td>
<td>0.9693*/39.8</td>
<td>0.9964*/38.8</td>
</tr>
<tr>
<td></td>
<td>Polynomial</td>
<td>0.9913*/40.5</td>
<td>0.9605/48.9</td>
<td>0.9964*/44.2</td>
</tr>
<tr>
<td></td>
<td>S-function</td>
<td>0.9826/41.2</td>
<td>0.9565/45.6</td>
<td>0.9964*/44.3</td>
</tr>
<tr>
<td>25</td>
<td>RBF</td>
<td>0.987/37.8</td>
<td>0.9693*/38.5</td>
<td>0.9927/38.7</td>
</tr>
<tr>
<td></td>
<td>Polynomial</td>
<td>0.9957*/36.9</td>
<td>0.9561/45.7</td>
<td>0.9964*/39.7</td>
</tr>
<tr>
<td></td>
<td>S-function</td>
<td>0.9826/35.4</td>
<td>0.9561/41</td>
<td>0.9927/43.8</td>
</tr>
<tr>
<td>20</td>
<td>RBF</td>
<td>0.9609/37.0</td>
<td>0.9649*/37.3</td>
<td>0.9745/37</td>
</tr>
<tr>
<td></td>
<td>Polynomial</td>
<td>0.9826*/33</td>
<td>0.9430/39.9</td>
<td>0.9891*/36.7</td>
</tr>
<tr>
<td></td>
<td>S-function</td>
<td>0.9652/38.2</td>
<td>0.9386/39.7</td>
<td>0.9708/38.2</td>
</tr>
<tr>
<td>15</td>
<td>RBF</td>
<td>0.9348/35.2</td>
<td>0.9167/36.8</td>
<td>0.9635/31.0</td>
</tr>
<tr>
<td></td>
<td>Polynomial</td>
<td>0.9870*/29.6</td>
<td>0.9254*/33.8</td>
<td>0.9672*/27.5</td>
</tr>
<tr>
<td></td>
<td>S-function</td>
<td>0.9652/34.3</td>
<td>0.9167/37.1</td>
<td>0.9518/34.2</td>
</tr>
<tr>
<td>10</td>
<td>RBF</td>
<td>0.8826/35.0</td>
<td>0.8509/35.9</td>
<td>0.8940/36.8</td>
</tr>
<tr>
<td></td>
<td>Polynomial</td>
<td>0.9609*/24.4</td>
<td>0.8728*/28.4</td>
<td>0.9015*/27.8</td>
</tr>
<tr>
<td></td>
<td>S-function</td>
<td>0.8565/35.1</td>
<td>0.8486/33.3</td>
<td>0.9708/35.7</td>
</tr>
</tbody>
</table>

Table 3: Performance comparisons among different kernel function SVMs
Face Detection and Recognition using Colour Sequential Images

- Human face detection accuracy: 97.2% under controlled lighting conditions.
- Human face (70 persons) recognition accuracy: 96.5% (with 20 eigenvectors) and 98.3% (with 30 eigenvectors)
- Time of human face detection and recognition: less than 3 seconds under the hardware environment Pentium III and software environment MS Windows 2000 and MS VC++ 6.0.

Further work will focus on the robustness of the system, to develop an algorithm that can detect a face and recognize it in different lighting conditions and against complex backgrounds.

ACKNOWLEDGEMENT

The authors wish to acknowledge that this work is supported by Fundamental Project of Shanghai under grant number 03DZ14015. The authors would like to thank Dr Yonggang Wang for his useful discussion.

REFERENCES


BIOPGRAPHICAL NOTES

Zhonglong Zheng received the BA degree in EE from the University of Petroleum, China in 1999. He is a PhD student at the Institute of Image Processing and Pattern Recognition, at Shanghai JiaoTong University, China. His current research interests include image analysis, video tracking, face detection and recognition, and data mining.

Zhonglong Zheng
Face Detection and Recognition using Colour Sequential Images

Jie Yang received the PhD degree in computer science from University of Hamburg, Germany in 1994. Dr Yang is now Professor and Vice-director of the Institute of Image Processing and Pattern Recognition at Shanghai JiaoTong University, China. His research interests include artificial intelligence, image processing, computer vision, data mining, and pattern recognition. He is the author of more than 80 scientific publications and books.

Yitan Zhu received his BA and MS degrees in pattern recognition and intelligent systems from Shanghai JiaoTong University, China. He is a PhD student at Alexandria Research Institute, Virginia Polytechnic Institute and State University, USA.