Pose-Invariant Face Recognition in Hyperspectral Images

Han Wang*
Department of Electrical Engineering and Computer Science
University of California, Irvine
Irvine, California 92697
Email: hwang5@uci.edu

Glenn Healey
Department of Electrical Engineering and Computer Science
University of California, Irvine
Irvine, California 92697
Email: ghealey@uci.edu

Abstract—Pose-invariant face recognition remains a challenging problem, especially when the pose change is large. Previous studies use either spatial or spectral information to address this problem. In this paper, we propose an algorithm that uses spatial and spectral information simultaneously to deal with large pose changes. We first learn 3D models from 2D images. We then use these 3D models to generate images in novel poses. Finally, we use spatial and spectral information to classify a test image. We demonstrate the effectiveness of the algorithm on a database of 200 subjects.

Keywords—Face recognition, hyperspectral, Gabor filter, principal-component analysis (PCA), gradient descent

I. INTRODUCTION

The performance of face recognition systems degrades significantly in the presence of large pose variations due to the fact that image differences caused by pose changes are often large. Many algorithms have been proposed to address the problem. These methods can be divided into two categories: 2D techniques and 3D methods. The former category uses only 2D images while the latter uses 3D face models. PDM [1], AAM [2] and TFA [3] are representatives from the first category. 2D methods often require images of different poses which may not be available in practice. Studies also show that 2D methods can tolerate pose variation up to a certain limit ($\pm 45^\circ$) and beyond that the performance deteriorates dramatically [4]. Due to the fact that human heads are inherently 3D objects, 3D model-based methods are becoming popular. Representatives from this category include 3DMM [5] and systems proposed by Jiang et al. [6], Ashana et al. [7] and Prabhu et al. [8]. However, these methods have not provided results in the presence of large pose change ($>60^\circ$). In this paper, we provide results on these pose changes as well. On the other hand, recent studies show that spectral discriminants also provide useful information and can be used for this purpose [9], [10], [11]. To make use of the flexibility of 3D models and spectral information, we propose an algorithm that learns personalized 3D models from 2D images and uses spatial and spectral information for recognition.

II. OVERVIEW OF THE ALGORITHM

The proposed face recognition algorithm consists of four stages. These stages are preprocessing, model training, feature extraction, and classification.

A. Preprocessing

Eye locations are first identified manually. Images are rotated such that the left eye and the right eye are at the same height in the rotated image. For profile images, forehead and mouth locations are identified. Images are rotated such that the angle between the line connecting the forehead and the mouth and the vertical axis is within a certain range. An example of original and rotated images is shown in Fig. 1.

B. Model Training

To learn a 3D face model, we use the 3D Basel Face Database [11] that is obtained from 200 face scans. The face model has 53490 vertices and we use 31133 vertices that comprise the face area. A face $S = (x_1, y_1, z_1, \ldots, x_n, y_n, z_n)^T$ is represented by $n$ vertices where $T$ is the transpose operation. It can also be represented using a basis according to

For any inquiries, contact this author.
where $\bar{S}$ is the mean shape, $P \in \mathbb{R}^{m \times n}$ is the shape basis and $m$ is the number of eigenvectors, and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)^T$ is the coefficient vector. In the study, we choose $m = 50$. We allow the face model to be scaled and translated according to

$$S = cS + T = c(S + Pa) + T$$

where $c$ is the scale ratio and $T$ is the translation.

To learn the parameter vector $(\alpha, c, T)$, we use correspondences between the 3D model and the 2D image. We aim to find the parameter set that minimizes the sum of the squared error between the two sets of correspondences given by

$$C = \sum_i ((c(S + Pa) + T - S_i')^2)$$

where $i$ denotes the $i_{th}$ correspondence and $S_i'$ is the correspondence in the image.

We use steepest descent with random step size [13] to find the optimal parameters. In the study, gallery images are frontal images. Since frontal images do not have depth information, we also use correspondences in the probe image (a different pose). In the study, we use 65 correspondences in the frontal image and 22 correspondences in the 45° image and 26 correspondences in the 90° images. Most correspondences are anthropometric facial points [14] which can be identified uniquely, e.g., corner of the eye. Some correspondences are identified by using neighboring correspondences, e.g., the midpoint of the two neighboring correspondences. An example of the correspondences in the frontal, 45° and 90° images is shown in Fig. 2.

After the parameter set is estimated, we reconstruct a 3D face model according to (2). The correspondences in the 3D model often do not coincide with those in the frontal image. We add displacements to correspondences in the face model to make the two overlap. We interpolate displacements for non-correspondence vertices and add them to the model. Therefore, we learn a personalized face model given by

$$S_i' = c_iS + Pa_i + T_i + \text{disp}$$

where $\text{disp}$ is the added displacement.

To learn texture information, we project the frontal image orthographically to the 3D model. For vertices that do not receive a texture assignment, their texture is interpolated.

C. Features

We use two types of features: spatial and spectral. Spatial features are extracted from 2D images. There are three kinds of spatial features: filtered images, correspondence PCA coefficients, and Gabor jets of correspondences.

For filtered image features, image regions centered around the eye and the mouth are used. An example of the image regions in 45° and 90° images is shown in Fig. 3 where white rectangles highlight the regions.

The image regions are filtered by Gabor filters. A Gabor filter is a sinusoidal function modulated by a Gaussian envelope given by

$$g(x, y) = a(x, y) \exp \left( -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right)$$

where

$$a(x, y) = \frac{1}{2\pi \sigma_x \sigma_y} e^{-\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right)}$$

is the Gaussian component and

$$c(x, y) = \exp(i(2\pi f (x \cos(\theta) + y \sin(\theta))))$$

is the sinusoidal component. The standard deviations $(\sigma_x, \sigma_y)$ define the size of the Gaussian envelope, $f$ is the center frequency in the frequency domain and $\theta$ is the center frequency orientation in the frequency domain. The filtered image magnitude is given by

$$y(x, y) = |g(x, y) \ast I(x, y)|$$

where $\ast$ denotes the convolution operation, $|$ is the magnitude operation, and $I(x, y)$ is the DC subtracted image region. To
alleviate the effect of misalignment in images, a large scale is used, i.e., \( x = 16 \). We use \( y = x/2 \) and \( f = 1/x \). Eight orientations are used, i.e., \( \theta = [0,1/8,\ldots,7/8]\pi \). This gives us 8 filtered images.

The second kind of spatial features are correspondence PCA coefficients. They are obtained by projecting correspondence coordinates \( \text{coor} \) to a basis given by

\[
\text{coor} = \text{coor} + B\beta + e
\]

where \( \beta \) is the coefficient, \( B \) is a basis, \( \text{coor} \) is the mean coordinate, and \( e \) is an error term. The basis \( B \) is obtained by applying PCA to correspondence coordinates across images of the same pose. For example, a basis for 45\(^\circ\) images is obtained by applying PCA to correspondence coordinates of all 45\(^\circ\) images. In this study, the number of eigenvectors used captures 95\% variation.

The third kind of spatial features are Gabor jets. They are obtained by applying Gabor filters at correspondences. A Gabor feature is defined as the magnitude of convolving a Gabor filter with an image point given by

\[
y = |g(x,y) * I(x,y)|
\]

where \( g(x,y) \) is the Gabor filter and \( I(x,y) \) is the DC subtracted image region centered around the point. The same Gabor filters used in extracting filtered image features are used to extract Gabor jet features. A Gabor jet is a vector \( y = (y_1, y_2, \ldots, y_8)^T \) consisting of the Gabor features obtained by applying all Gabor filters to a correspondence.

Spectral features are represented by a 31-dimensional spectral mean vector extracted from a tissue sample. Three tissues are used. They are cheek, chin and hair. Sample locations are determined by moving away a fixed distance from a known location. For example, the left cheek sample in the frontal image is 60 pixels below the left eye and 15 pixels left of the left eye. Chin and hair samples are found similarly by using the mouth and forehead as references. We use size 11x11 for cheek and chin samples and 5x5 for hair sample. The spectral feature vector is obtained by averaging the reflectance of the pixels in the sample and normalizing the vector by its length.

\[
R(\lambda) = \frac{1}{n} \sum_{x,y} R(x,y,\lambda) / \left| \sum_{x,y} R(x,y,\lambda) \right|
\]

where \( R(x,y,\lambda) \) is the reflectance at location \( (x,y) \) and wavelength \( \lambda \) and \( n \) is the number of pixels in the sample.

Fig. 4 (a) gives an example of the three tissue samples highlighted by the squares. The spectral vectors for the three samples are shown to the right.

Spectral features have spatial variability depending on the location. To reduce the dependence on the location, we use a larger window to extract a set of spectral vectors. The window width and height is 3 times the size of the sample region and the spectral vector is extracted in regions every two pixels apart.

D. Classification

For each probe image, we use its correspondences and the correspondences of each gallery image to generate a 3D face model. We rotate the 3D model according to the pose of the probe. We then orthographically project the model to an image plane to generate a virtual image. An example of a frontal image, two rotated images and the virtual images of a subject is shown in Fig. 5.

The virtual image is then used as the face image of the subject used to generate the model. Spatial features are then extracted from the virtual image and the probe respectively. For filtered image features, we use the Euclidian distance to measure the similarity between the two filtered images given by

\[
\text{Sim}(I_1, I_2) = \frac{1}{nm} \sum_{x,y} |I_1(x,y) - I_2(x,y)|
\]
\[ d'_j = \frac{\sum_{i,j} (y'_{i,j}(x,y) - y_{i,j}(x,y))^2}{n'_j} \]  

(12)

where \( y_{i,j}(x,y) \) and \( y_{i,j}(x,y) \) are the filtered virtual and probe images using filter \( j \) respectively, \( t \) represents the image region (eye or mouth), and \( n'_j \) is the number of pixels in the region. To allow misalignment between \( y_{i,j}(x,y) \) and \( y_{i,j}(x,y) \), we allow one to be shifted between \([-12, 12]\) pixels with respect to the other and the one that minimizes (12) is chosen to define the distance.

For correspondence PCA features, we project correspondence coordinates in the virtual and probe images to the basis of the probe pose to obtain their coefficients according to (9). We use the Euclidian distance to measure the similarity between the two sets of coefficients given by

\[ d_c = \sum (\beta_i - \beta_j)^2 \]  

(13)

where \( i \) denotes the \( i_{th} \) coefficient.

We also extract Gabor jet features from the virtual and probe images according to (10). We use the Euclidian distance to measure the similarity between two Gabor jets given by

\[ d_g = \sum_{i,j} (y'_{i,j} - y_{i,j})^2 \]  

(14)

where \( i \) represents the \( i_{th} \) correspondence and \( j \) the \( j_{th} \) Gabor feature.

For spectral features, for each tissue type \( t \), i.e., cheek, chin and hair, we extract a set of spectral vectors from the hyperspectral probe and the original hyperspectral gallery respectively. We use the Mahalanobis distance and the distance between the two sets of spectral vectors is defined as the smallest distance between the two sets given by

\[ d_s = \min_{k \in \text{set}, l \in \text{set}} (\Sigma^{-1}(\bar{R}_{s,k} - \bar{R}_{s,l})) \]  

(15)

where \( g \) represents the gallery image and \( m \) is the number of spectral vectors in each set. \( \Sigma \) is the covariance matrix of the spectral vectors for tissue type \( t \) and is approximated by a diagonal matrix with elements corresponding to the variance of the spectral features in the database.

The total distance is a weighted average of all of the distance metrics given by

\[ d = \sum_i c'_i d'_i + \sum_i c'_d d'_d + c_s d_s + c_g d_g \]  

(16)

where the weights are the reciprocal of the standard deviation of the smallest distances of all subjects in each distance metric respectively.

We used a face database of 200 subjects for our experiments [9]. All images have 31 bands with center wavelengths separated by 0.01 \( \mu \text{m} \) over the near-infrared (NIR) (0.7 \( \mu \text{m}-1.0 \mu \text{m} \)). The spatial resolution is 494x468 pixels. We consider five images for each subject. Image \( f_g \) is a frontal image, image \( f_{l1} \) is a left 45° image, image \( f_{r1} \) is a right 45° image, image \( f_{l2} \) is a left 90° image, and image \( f_{r2} \) is a right 90° image. An example of the five images is shown in Fig. 6.

We use frontal images as the gallery and the other four images as probes. We use all subjects as test subjects. The experiment follows the closed universe model.

The classification rates using different features for the four poses are summarized in Table I where spatial means using all three spatial features. To compare with other algorithms, we used the EBGM method provided by the CSU Face Identification Evaluation System [15], [16], [17] and included the result in Table I.

**III. EXPERIMENTS**

We used a face database of 200 subjects for our experiments [9]. All images have 31 bands with center wavelengths separated by 0.01 \( \mu \text{m} \) over the near-infrared (NIR) (0.7 \( \mu \text{m}-1.0 \mu \text{m} \)). The spatial resolution is 494x468 pixels. We consider five images for each subject. Image \( f_g \) is a frontal image, image \( f_{l1} \) is a left 45° image, image \( f_{r1} \) is a right 45° image, image \( f_{l2} \) is a left 90° image, and image \( f_{r2} \) is a right 90° image. An example of the five images is shown in Fig. 6. We use frontal images as the gallery and the other four images as probes. We use all subjects as test subjects. The experiment follows the closed universe model.

The classification rates using different features for the four poses are summarized in Table I where spatial means using all three spatial features. To compare with other algorithms, we used the EBGM method provided by the CSU Face Identification Evaluation System [15], [16], [17] and included the result in Table I.

**TABLE I. CLASSIFICATION RESULTS FOR THE DATABASE**

<table>
<thead>
<tr>
<th>Method</th>
<th>Pose</th>
<th>45° left</th>
<th>45° right</th>
<th>90° left</th>
<th>90° right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td></td>
<td>0.92</td>
<td>0.92</td>
<td>0.13</td>
<td>0.1</td>
</tr>
<tr>
<td>Coefficient</td>
<td></td>
<td>0.09</td>
<td>0.09</td>
<td>0.17</td>
<td>0.09</td>
</tr>
<tr>
<td>Gabor jet</td>
<td></td>
<td>0.37</td>
<td>0.47</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>Spatial</td>
<td></td>
<td>0.9</td>
<td>0.91</td>
<td>0.28</td>
<td>0.24</td>
</tr>
<tr>
<td>Spectral</td>
<td></td>
<td>0.85</td>
<td>0.78</td>
<td>0.69</td>
<td>0.66</td>
</tr>
<tr>
<td>Spectral+spatial</td>
<td></td>
<td>0.99</td>
<td>0.97</td>
<td>0.84</td>
<td>0.77</td>
</tr>
<tr>
<td>EBGM</td>
<td></td>
<td>0.87</td>
<td>0.85</td>
<td>0.02</td>
<td>0.03</td>
</tr>
</tbody>
</table>

When the probe pose is \( \pm 45^\circ \), the EBGM method, which only uses 2D information, gives more than 85% accuracy, suggesting that at this point spatial information in the frontal pose can still provide useful information. By
using a 3D model, performance can be improved further as shown in the spatial features. When the probe pose changes to $\pm 90^\circ$, both spatial and EBGM methods are greatly affected. With a pose change so large, the EBGM method degrades significantly. This suggests that frontal information alone is not enough when the two poses are very different. Nevertheless, the proposed 3D method provides additional information that is helpful in this case as shown by the spatial features. On the other hand, spectral features degrade moderately across pose. This is consistent with previous studies [9], [10], [11] that show that spectral features are less affected by pose changes. No matter what pose it is, the combined method using spatial and spectral features outperforms using either one alone, suggesting spatial and spectral features provide distinct information. When the probe pose is $\pm 90^\circ$, the combined method outperforms the EBGB method by more than 70%.

The cumulative match score for the four poses is shown in Fig. 7 where rank N means the correct identity is within the top N candidates selected by the algorithm.

For the two $45^\circ$ poses, the score of the combined method continues to improve as the rank increases. For the two $90^\circ$ poses, the score picks up quickly and reaches 95% at rank 6 suggesting that the correct match is more similar to the probe than most candidates if it is misclassified in the first round.

Since frontal images provide a significant amount of information when the pose changes to $\pm 45^\circ$, we simplified the algorithm for this case. To estimate the 3D model, we only use correspondences from the frontal image. In classification, only filtered image features are used. The classification results on the two $45^\circ$ poses are shown in Table 2. For convenience, results obtained using the EBGM method are also included.

![Fig. 7. Cumulative match score for (a) left $45^\circ$ image (b) right $45^\circ$ image (c) left $90^\circ$ image (d) right $90^\circ$ image](image-url)
The result using filtered image features is comparable to the previous case where a personalized 3D model is used. It also outperforms the EBGM method by 5%, showing that a generic 3D model also helps in this case. The combined method is better than using either spatial or spectral features alone and outperforms the EBGM method by 10%. The cumulative match score using this approach is shown in Fig. 8.

<table>
<thead>
<tr>
<th>Method</th>
<th>Pose</th>
<th>45° left</th>
<th>45° right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td></td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Spectral+spatial</td>
<td></td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>EBGM</td>
<td></td>
<td>0.87</td>
<td>0.85</td>
</tr>
</tbody>
</table>

The cumulative score is similar to the previous case suggesting that a generic 3D model is sufficient when the pose change is not larger than 45°.

I. CONCLUSION AND FUTURE WORK

We have presented an algorithm for pose-invariant face recognition in hyperspectral images which uses both spatial and spectral information. We learned 3D face models from 2D images and used the models to generate virtual images at a different pose. Spatial features were extracted from the virtual images and were used with spectral features to recognize a test image. Compared to other related work, the proposed method provides a novel way to reconstruct 3D face models from 2D images and shows the effectiveness of using spatial and spectral information simultaneously. Future work could include using an advanced classifier to make greater use of the virtual images and integrating pose estimation into the framework.

ACKNOWLEDGEMENT

This research has been supported by an Army Research Office Grant.

REFERENCES


