Pose-Robust Face Recognition Based on Texture Mapping

Kwang Ho An and Myung Jin Chung, Member, IEEE

Abstract—A human face provides a variety of different communicative functions such as identification, the perception of emotional expression, and lip-reading. Many applications in robotics require recognizing a human face. A face recognition system should be able to deal with various changes in face images, such as pose, illumination, and expression, among which pose variation is the most difficult one to deal with. For this reason, face registration is the key of face recognition. If we can register face images into frontal views, the recognition task would be much easier. To align a face image into a canonical frontal view, we need to know the pose information of a human head. A human head can be approximately modeled as a 3D texture mapped ellipsoid. Then, any face image can be considered as a 2D image projection of a 3D ellipsoid at a certain pose. In this paper, both training and test face images are back projected to the surface of a 3D ellipsoid according to their estimated poses and registered into canonical frontal view images. Then, simple and efficient frontal face recognition can be carried out in the texture map domain instead of the original input image domain. To evaluate the feasibility of the proposed approach, several recognition experiments are conducted using subspace-based face recognition methods such as PCA, PCA+LAD, and DCV. By conducting experiments on our laboratory database and the Yale Face Database B, we show that the proposed algorithm provides good performance even when large out-of-plane rotations occur.

I. INTRODUCTION

FACE recognition has been studied by many researchers in the past two decades [1], [3], [8]. A human face provides a variety of different communicative functions such as identification, the perception of emotional expression, and lip-reading. Many applications in robotics require recognizing a human face. Vision-based face recognition has several advantages over other biometric technologies. It is natural, non-intrusive and easy to use. Continuously finding the position of the user’s face and recognizing the user may be necessary for the computer to accomplish tasks such as surveillance, user-dependent service, and identity authentication. A face as a three-dimensional object subject to varying illumination, pose, and expression is to be identified based on its two-dimensional image in 2D face recognition. A face recognition system generally consists of four major modules as depicted in Fig. 1. A face recognition system should be able to deal with various changes in face images, such as pose, illumination, and expression, among which pose variation is the most difficult one to deal with [15], [16]. For these reasons, face registration is the key of face recognition. Many approaches have been proposed for pose-robust face recognition [2], [5], [6]. Approaches for addressing pose variation can be largely classified into two categories. The first type of these approaches is often called multi-view face recognition [2], [9], [10]. Multi-view face recognition is a simple extension of frontal face recognition. It treats the whole face image under a certain pose as one vector in a high-dimensional sample space. And the training is done using multi-view face images and a test image is assumed to be matched to one of the existing head poses. Generally, multi-view based approaches should have view-specific classifiers. Therefore, the training and recognition processes are even more time consuming. The second type of approaches is face recognition across pose [5], [6], [11]. It uses a canonical view for face recognition. This method needs a face alignment process to generate a novel frontal view image. Therefore, various well-known frontal face recognition methods can be easily applied to this type of approaches. In this paper, we adopt the latter approach. If we can register face images into frontal views, the recognition task would be much easier. We can approximately model a human face as a 3D ellipsoid. Then, any face image can be considered as a 2D image projection of a 3D ellipsoid at a certain pose. In other words, both training and test face images are back projected to the surface of a 3D ellipsoid according to their estimated poses and registered into canonical frontal view images. Then, face recognition can be carried out in the texture map domain instead of the original image domain, which is approximately linear-separable space. To evaluate the feasibility of the proposed approach, several face recognition experiments are conducted using linear subspace-based face recognition methods such as PCA, PCA+LAD, and DCV [3], [7], [12]. By conducting experiments on our laboratory database and the Yale Face Database B, we show that the proposed algorithm provides good recognition performance even when large out-of-plane rotations occur.

![Face Recognition Flow Diagram](image-url)
In this paper, we propose a pose-robust face recognition algorithm. In section 2, we first introduce how to model a human face as a 3D ellipsoid. Section 3 describes how to generate a texture map from a face image by using a simple 3D ellipsoid model. In section 4, we review three traditional subspace-based face recognition methods such as PCA, PCA+LDA and DCV, which are used to evaluate our proposed approach. In section 5, we show our various experimental results to verify the feasibility of our method. Finally, we conclude the paper in section 6.

II. 3D FACE MODELING

A human head has non-planar geometry. Also, it has curved surfaces. Therefore, to build a 3D model of a human head simply, a cylinder or an ellipsoid has been used. Among them, we adopt a 3D ellipsoid model to represent a human head since a cylinder model does not represent vertically curved surfaces compared with a 3D ellipsoid model. A 3D ellipsoid itself is parameterized by the lengths of its major axes. We assume that the width \( r_x \) is 1. Thus we only need to determine the ratios between the width \( r_x \) to the height \( r_y \) and the width to the depth \( r_z \). In our approach, we statistically obtain these ratios from the sample data of human heads to represent curved surfaces of a human head more generally. Also, we only use the partial regions of a human head with a range of \( 60^\circ \leq \alpha \leq 135^\circ \) and \( -90^\circ \leq \beta \leq 90^\circ \) to express a human face more precisely and exclude disturbing regions such as hairs and background outliers. Fig. 2 illustrates how to model a human head as a 3D ellipsoid. The origin of an object coordinate is placed at the center of a human head. And the frontal face looks at the positive \( Z \) axis of an object coordinate.

III. TEXTURE MAPPING

If we compare two face images of the same subject taken under two different poses, the pixel-wise difference is relatively big because these two images are not aligned with respect to each other. Therefore, the traditional face recognition methods based on frontal view are not working well. Face image registration is needed for this reason. If we can build a 3D model of a human head and estimate a current pose, we can obtain a stabilized (frontal) view image. And finally, we can implement a pose-robust face recognition system. First, we assume that a human head is a 3D ellipsoid with radiiuses \( r_x, r_y, \) and \( r_z \) as explained above. Also, face images are captured with the perspective projection camera model. Finally, the camera is intrinsically calibrated. Under these assumptions, we can find the relation between an arbitrary 2D face image and its stabilized texture map if we only know the pose information. \( \mathbf{R} \) and \( \mathbf{T} \) in Fig. 3 define the 3D pose of a human head. We can manually estimate the current pose of a human head by adjusting the pose parameters. And also, we use a 3D model-based image registration technique to estimate the current pose of a human head. The procedure of texture mapping is as follows. First, a pixel in the texture map corresponds to one point on the surface of an ellipsoid because we obtain the texture map by projecting the surface points onto \( X_O Y_O \) plane. Second, we can estimate the current pose of a 3D ellipsoid manually or by an image registration technique. Third, we can project all surface points of the ellipsoid to the input image plane using the perspective projection model. Finally, we can find out the complete relationship between an arbitrary input face image and its corresponding texture map and obtain a stabilized (frontal) view image. However, there might be missing pixels in the stabilized texture map, which correspond to invisible regions from the camera. Fortunately, we know that a human face has a symmetric property around \( Y_O \) axis. Therefore, we can also generate a mirrored texture map by a simple mirror operation. By doing so, we can make
up for the missing pixels and improve the recognition performance. Fig. 3 shows the geometrical relationship between a camera coordinate frame and an object coordinate frame. Let \( \begin{bmatrix} X_O & Y_O & Z_O \end{bmatrix}^T \) be the coordinates of \( \mathbf{P}_O \) in the object reference frame and \( \begin{bmatrix} X_C & Y_C & Z_C \end{bmatrix}^T \) be the coordinates of \( \mathbf{P}_C \) corresponding to the point \( \mathbf{P}_O \). Then, the relationship between two coordinate frames can be represented by (1). \( \mathbf{R} \) represents the 3\( \times \)3 rotation matrix and \( \mathbf{T} \) is the 3D translation vector.

\[
\mathbf{P}_C = \mathbf{R}\mathbf{P}_O + \mathbf{T} \quad (1)
\]

\[
\mathbf{P}_C = \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix}, \quad \mathbf{P}_O = \begin{bmatrix} X_O \\ Y_O \\ Z_O \end{bmatrix}, \quad (2)
\]

where \( (X_O,Y_O,Z_O) \) can be defined as (5) and (6) by using a 3D ellipsoid model.

\[
\mathbf{R} = \text{Rot}(z, \phi)\text{Rot}(y, \theta)\text{Rot}(x, \psi) = \begin{bmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & -\sin \psi \\ 0 & \sin \psi & \cos \psi \end{bmatrix} \quad (3)
\]

\[
\mathbf{T} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}^T \quad (4)
\]

\[
X_O = r_z \sin \alpha \sin \beta, \quad Y_O = r_z \cos \alpha, \quad Z_O = r_z \sin \alpha \cos \beta \quad (5)
\]

\[
60^\circ \leq \alpha \leq 135^\circ, \quad -90^\circ \leq \beta \leq 90^\circ \quad (6)
\]

Also, let \( \mathbf{p} \) be the image coordinates of \( \mathbf{P}_C \). Then, it can be defined like below under the perspective projection camera model.

\[
\mathbf{p} = \begin{bmatrix} x_p \\ y_p \end{bmatrix} \quad (7)
\]

\[
x_p = f \frac{X_C}{Z_C}, \quad y_p = f \frac{Y_C}{Z_C} \quad (8)
\]

Fig. 4(a) illustrates how to obtain a texture map when given an input face image and its pose parameters. One point on the 3D ellipsoid corresponds to a pixel on the input image under an estimated pose.
Therefore, we can generate a stabilized (frontal) view image shown in the left of Fig. 4(b) by projecting all points \( P_O \) which belong to the range of (6) onto the \( X_O Y_O \) plane. As I mentioned above, a human face is symmetric around \( Y_O \) axis. Therefore, we can improve the recognition performance by using this helpful property when there are missing pixels in the stabilized view image due to the camera’s viewing direction. The right image of Fig. 4(b) is a mirrored image of (a) by a simple mirror operation. Fig. 5 shows the sample face images of 10 subjects from the Yale Face Database B. Fig. 6 shows the stabilized view images and their mirrored images obtained by texture mapping.

IV. FACE RECOGNITION

In this section, we briefly review the traditional linear subspace-based face recognition methods such as Principal Component Analysis (PCA), Fisherface method (PCA+LDA), and Discriminative Common Vector (DCV), which are used for verifying the feasibility of our proposed face recognition framework.

A. PCA

A technique has been commonly used for dimension reduction in computer vision, especially in face recognition, is Principal Component Analysis (PCA) [3], [4]. Many methods have been proposed for face recognition within the last two decades. Among these methods, appearance-based methods operate directly on the images and process the images as two-dimensional holistic patterns. In these approaches, a two-dimensional image of size \( w \times h \) pixels is represented by a vector in a \( wh \)-dimensional space. However, the dimension of this image space is typically very high. Since face images are highly correlated, they can be represented in a low-dimensional subspace without losing a significant amount of information. PCA is one of the most famous methods for finding such a lower-dimensional subspace. The key idea of PCA is to find the best set of principal components that maximize the total scatter across all sample images. The objective function of PCA is as follows.

\[
W' = \arg \max_{W} \left| W^T S_T W \right|, \tag{9}
\]

where \( S_T \) is the total scatter matrix of the training samples and \( W \) is the matrix whose columns are the orthonormal projection vectors. Generally, the eigenvectors of \( W \) are called the eigenfaces.

B. PCA+LDA

Linear Discriminant Analysis (LDA) has been one of the most popular techniques employed in the face recognition [7], [8]. The basic idea of LDA is to calculate the optimal discriminant vectors so that the ratio of the between-class and within-class scatter is maximized. In other words, LDA is used to seek an optimal projection matrix, from the original sample space to a lower dimensional space, which maximizes the between-class scatter matrix while minimizing the within-class scatter matrix.

\[
W^* = \arg \max_{W} \left| \frac{W^T S_B W}{W^T S_W W} \right| \quad \tag{10}
\]

where \( S_B \) is the between-class scatter matrix and \( S_W \) is the within-class scatter matrix. The optimal solution can be obtained when the columns of the projection matrix \( W \) are the eigenvectors of \( S_W^{-1} S_B \). Unfortunately, \( S_W \) is mostly singular in many applications of pattern recognition since the dimension of the sample space is much larger than the number of samples in the training database. This is called a small sample size problem. To solve this problem, various methods have been proposed in the last decade. Swets and Weng proposed the Fisherface method [7]. They use PCA for dimension reduction to make the within-class scatter matrix \( S_W \) nonsingular. And then, LDA is applied to obtain the optimal projection matrix \( W \). In this paper, we call it PCA+LDA method.

C. DCV

The key idea of Discriminant Common Vector (DCV) approach is that it extracts the common properties of training samples by eliminating the differences in each class [12], [13]. The null space of \( S_W \) contains considerable discriminative information for face recognition. However, PCA+LDA method discards this information by removing the null space of \( S_W \). To use such discriminative information, all the samples are firstly projected onto the null space of \( S_W \), where the within-class scatter matrix is zero, and then the optimal discriminant vectors are those vectors that can maximize the between-class scatter matrix of common vectors. A common vector for each class is obtained by removing all the features that are in the direction of the eigenvectors corresponding to the nonzero eigenvalues of the scatter matrix of its own class.

\[
W^* = \arg \max_{W} \left| \frac{W^T S_B W}{W^T S_W W} \right| \quad \tag{11}
\]

For obtaining the optimal projection matrix, vectors that span the null space of \( S_W \) must first be computed. Since the dimension of this null space can be very high, it is computationally expensive. A more efficient way is to use the orthogonal complement of the null space of \( S_W \), which typically is a significant lower-dimensional space.
V. EXPERIMENTAL RESULTS

We evaluated our proposed algorithm on the Yale Face Database B and our laboratory database with the traditional subspace-based face recognition methods such as PCA, PCA+LDA, and DCV. The Yale Face Database B is to allow systematic testing of face recognition methods under large variations in illumination and pose [14]. The database contains 5760 single light source images of 10 subjects under 576 viewing conditions (9 poses × 64 illumination conditions). Images of 10 subjects were captured under 64 lighting conditions in 9 poses. For the Yale Face Database B, we manually estimate the poses of subjects and generate stabilized and mirrored face images. In the first experiment, we used one frontal face image for the training and the other 8 face images with pose variations for recognition test from each class. Face recognition was performed with the traditional PCA. Face recognition results are shown in Table I. In the second and third experiments, we used two frontal face images for the training and tested 16 face images of 8 poses with similar illumination in each class. The recognition results are shown in Table II and III respectively. In the second experiment, we used PCA+LDA method. And DCV was used in the third experiment. Our laboratory database consists of 372 face images of 2 subjects with large motion. And illumination is not varying significantly. The ranges of motion in our laboratory database are shown in Table V. To build our laboratory database, we used a 3D model-based face tracking algorithm using optical flow-based technique for image sequences of 2 subjects as shown in Fig. 7.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>RECOGNITION RESULTS WITH PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unregistered</td>
</tr>
<tr>
<td>Recognition Rate</td>
<td>55%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>RECOGNITION RESULTS WITH PCA+LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unregistered</td>
</tr>
<tr>
<td>Recognition Rate</td>
<td>58.75%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>RECOGNITION RESULTS WITH DCV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unregistered</td>
</tr>
<tr>
<td>Recognition Rate</td>
<td>51.875%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE IV</th>
<th>RECOGNITION RESULTS WITH DCV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unregistered</td>
</tr>
<tr>
<td>Recognition Rate</td>
<td>63.0882%</td>
</tr>
</tbody>
</table>

In this experiment, the initial pose parameters for the head were assumed to be known. From the estimates for $R$ and $T$ by the tracking algorithm, we can generate stabilized (frontal) view images and mirrored images. Fig. 7 shows the results of 3D model-based face tracking. Fig. 8(a) and (b) are stabilized and mirrored face images of 2 subjects.

![Fig. 7. 3D model-based face tracking results with large motion.](image-url)
Fig. 8. Stabilized and mirrored face images from our laboratory database. Left two columns are stabilized (frontal) view images. And right two columns represent their mirrored images.

<table>
<thead>
<tr>
<th>TABLE V</th>
</tr>
</thead>
<tbody>
<tr>
<td>THE RANGE OF MOTION IN OUR LABORATORY DATABASE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Degree</td>
</tr>
</tbody>
</table>

And Table IV shows the results of face recognition on our laboratory database with DCV. In all experiments, we can verify that face registration (alignment) is helpful to improve the recognition performance and also the performance of face recognition with mirrored face images is much better. This is because that not all pixels on the stabilized view images can be visible from the camera under large motion (out-of-plane rotation and translation). Therefore, such regions are considered as missing pixels and set the intensities to be zeros in the stabilized view image. On the other hand, for mirrored face images, we can make up for the missing pixels with the symmetric property of a human face as shown in Fig. 8.

VI. CONCLUSIONS

In this paper, we proposed a pose-robust face recognition algorithm. A human head can be approximately modeled as a 3D texture mapped ellipsoid. Therefore, any face image can be back projected to the surface of a 3D ellipsoid according to its estimated pose and registered into canonical frontal view. Then, various well-known frontal face recognition methods can be easily applied to the stabilized view images. We have shown the feasibility of our proposed approach by conducting several experiments with traditional face recognition methods on the Yale Face Database B and our laboratory database and verified that our proposed method improved the performance of face recognition with pose variations.

REFERENCES