

Pose-Invariant Recognition of Faces at Unknown Aspect Views

Ashit Talukder and David Casasent

Dept. of ECE, Carnegie Mellon University, Pittsburgh, PA 15213

ashit@ece.cmu.edu, casasant@ece.cmu.edu

Abstract

A new technique is discussed to recognize human faces under varying aspect views (pose). We first estimate the pose of an unknown human face from a 2-D gray-scale image and then transform the unknown face to a reference pose using a feature extraction procedure. A different set of features for discriminating between different individuals are then extracted from these reconstructed faces for recognition. The feature extraction scheme used is known as the maximum representation and discrimination feature (MRDF) method. The advantage of our procedure is that it inherently removes distortions due to pose variations, and therefore requires only single training and/or test face images, which could be at different aspect views. For transformation, it does not require the face to be in the database during training. For recognition, only one aspect view at any pose is necessary.

1 Introduction

Several face recognition methods have been suggested in the recent past due to an increasing demand for security, surveillance, human-computer interaction, video compression for video-conferencing, and computer graphics (facial animation etc.). Prior face recognition algorithms used the concept of eigen-faces[1], joint transform correlators[2], and deformable intensity surface models[3]. Eigen-based techniques for pose-invariant face recognition have involved modular eigenspaces[4] in which one eigen-feature space is used for a range of aspect views of different people, and different view-based eigenspaces are used for different pose ranges. It has been shown[3] that global KL features (as used in modular eigenspaces) are not useful for discrimination if the covariance matrices of different classes are similar (this is likely to occur in this current problem in which the classes correspond to different orientation ranges for many faces). A method has been suggested to build a mathematical correspondence model between any face at 2 different

poses using the coordinates of facial parts such as the nose, mouth, etc. to generate synthetic face images from real gray-scale images[6]. Our method to achieve this does not need local facial features and is thus preferable.

Most of these face recognition techniques assume the presence of training and test faces at nearly the same frontal view or training faces at several orientations. We assume the presence of only a single training and test view, both of which are at different orientations (defined as side-to-side head movements) as well as some up/down motion). In this paper, we suggest a new method to estimate the pose of an unknown face, synthetically generate an image of the face at a reference pose, and then carry out recognition. A block diagram of our approach is shown in Fig. 1. Once the system is trained, this procedure is applied to any previously unseen face image at an unknown aspect view to remove distortions due to pose variations in the training or test set. At each of the 3 stages in our face recognition algorithm (shown in the 3 boxes in Fig. 1), we use new features designed by us that have been shown in prior-work[5, 7] to outperform other standard feature extraction techniques such as Fisher linear discriminant, Fukunaga-Koontz, Karhunen-Loeve (KL), and orthogonal discriminant vectors. The main underlying method in our research involves the nonlinear MRDF feature extraction procedure. The MRDF[7] has been successfully used to extract nonlinear features for simultaneous estimation of the pose and class of various similar objects[5]. This technique extracts nonlinear information from images without significant increased computational complexity compared to linear approaches, and provides closed-form solutions for the nonlinear transform to use. We discuss our approach for automated nonlinear feature extraction concept for pose estimation (Sect. 2), the mathematical formulation to find the correspondence between a face at two poses using our features (Sect. 3), and face recognition with the pose-transformed faces (Sect. 4). Results are presented in Sect. 5.

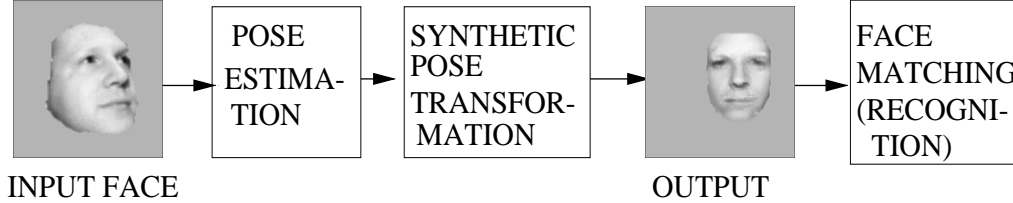


Figure 1: Block diagram of our approach for pose-invariant face recognition.

2 Facial Pose Estimation

We consider faces at each pose to be a single object class, and the goal is to extract discriminatory MRDF features such that faces at different poses are well separated in this feature space.

2.1 Discriminatory Nonlinear MRDF Feature Extraction From Images For Facial Pose Estimation

For discrimination, our objective is to find a nonlinear transformation (or a set of nonlinear transforms) on a class vector \underline{x} that optimally discriminates between classes in a reduced dimensionality space. We consider nonlinear transforms ϕ that are polynomial mappings of the input face images without any cross-product terms. Therefore, the nonlinear transformation ϕ we use for an image, $\underline{x} = [x_1 \dots x_N]^T$, is $y = \phi^T \underline{x}_p = \sum_{i=1}^N \phi_i x_i^p$, where $\underline{x}_p = [x_1^p \ x_2^p \ \dots \ x_N^p]^T$; p is the degree or order of nonlinearity. The computation time for this operation is $\mathcal{O}(N)$, similar to that of a correlation operation since it does not involve cross-product terms. For simplicity, we assume two classes and denote image samples from the two object classes as vectors \underline{x}_1 and \underline{x}_2 . The nonlinear data for class 1 images are then $\underline{x}_{1,p} = [x_{11}^p \ x_{12}^p \ \dots \ x_{1N}^p]^T$ and for class 2 images are $\underline{x}_{2,p} = [x_{21}^p \ x_{22}^p \ \dots \ x_{2N}^p]^T$. The ϕ_n are constrained to be orthogonal, implying that the output features y_n are uncorrelated. The separation measure to be maximized is

$$E_D = \sum_{n=1}^M E_{D,n} = \sum_{n=1}^M \frac{\phi_n^T \underline{R}_{12,p} \phi_n}{\phi_n^T (\underline{C}_{1,p} + \underline{C}_{2,p}) \phi_n}, \quad (1)$$

ie. we maximize $\underline{R}_{12,p} = \sum_{q=1}^Q \sum_{s=1}^S (\underline{x}_{1,p} - \underline{x}_{2,p})(\underline{x}_{1,p} - \underline{x}_{2,p})^T$, the mean squared separation between all class 1 and 2 projections on each basis vector ϕ_n , while minimizing the spread of each class $\underline{C}_{1,p}$ and $\underline{C}_{2,p}$ in feature space. For a given p , the nonlinear MRDF transformation coefficients correspond to the dominant eigen-vectors of the matrix $[\underline{C}_{1,p} + \underline{C}_{2,p}]^{-1}[\underline{R}_{12,p}]$. When there are L image object classes (or L possible aspect views for a human face), the nonlinear transformation that best separates all L classes (face images at different aspect views) are

the eigen-vectors of $[\sum_{i=1}^L \underline{C}_{i,p}]^{-1} [\sum_{i=1}^{L-1} \sum_{j=i+1}^L \underline{R}_{ij,p}]$, where $\underline{R}_{ij,p}$ is a higher-order correlation matrix that measures the separation between classes i and j , and $\underline{C}_{i,p}$ is the higher-order covariance matrix for class i .

The value for p for our pose estimation problem (and for face recognition) can easily be determined from training data[3] or by evaluating the pose-estimation (face classification) performance on a validation set. This approach outperforms standard ones since it computes ϕ_n that maximize the differences between all samples; standard methods only maximize the differences between class means or cannot easily handle more than 2 classes of data.

2.2 Modified k-NN Classifier for Facial Pose Estimation in Feature Space

We used a new modified k nearest neighbor classifier[8] for pose estimation. For each test sample at an unknown pose, we first extract discriminatory nonlinear MRDF features. We compute the average distance $d_{k_{avg}}^i$ of each test sample in MRDF feature space to the closest k prototype samples of each pose for all possible L pose-classes. The test sample is assigned to the pose-class with the closest average distance, i.e., the test sample is estimated to have pose n if $d_{k_{avg}}^n < d_{k_{avg}}^i \ \forall i = 1, 2, \dots, L, i \neq n$.

3 Facial Pose Transformation

The general mathematical formulation we use for pose transformation is similar to the one developed by Vetter and Poggio[6, 9] but uses our very different nonlinear MRDF features for representation.

3.1 Nonlinear MRDF Features for Representation of Human Faces

When it is desired to transform a face from one pose to another, it is necessary to have a good representation for a face in feature space[10]; for intra-class separation (of samples in the same class), the spread of the face image data at each pose (the projections onto the feature space) should be large. The representation measure we use in the MRDF is similar to the one used

in principal component analysis. Our goal is to extract orthonormal (uncorrelated) nonlinear features that minimize the mean-squared error between the original and reconstructed images for the sample vectors $\{\underline{x}_{1,p}\}$ in one class (as in PCA). We create the augmented data vectors, $\{\underline{x}_{1,p}\}$ for the class; these contain higher-order polynomial terms. For representation, the representation criterion to be maximized is

$$E_R = \sum_{m=1}^M \phi_m^T \underline{C}_{1,p} \phi_m, \quad (2)$$

where $\underline{C}_{1,p}$ is the higher-order covariance matrix of the augmented vectors $\{\underline{x}_{1,p}\}$. The M best nonlinear MRDF basis functions for representation ϕ_m are the M dominant eigenvectors of $\underline{C}_{1,p}$. In this initial research, we used only linear MRDFs features for representation, i.e. $p = 1$.

3.2 Pose Transformation Using MRDF Representation Features

We use linear MRDF features for representation (eigenfeatures) to generate synthetic face images at various orientations. Any gray-scale image \underline{x} of an unknown test face at the pose v can be optimally (in a least mean-squared-error sense) written as a linear combination of I linear MRDF transform vectors for representation ϕ_i at that pose v $\underline{x} \simeq \sum_{i=1}^I y_i \phi_i$ (as discussed in Sect. 3.1). The projection coefficients y_i for different test individuals are expected to be different and the salient characteristics for each person are preserved in the linear MRDF transform. The ϕ_i at pose v are determined from a set of training images of individuals.

Our goal is to generate an image of the unknown and previously unseen input face image \underline{x}' at the reference pose r , given its image \underline{x} at an input pose v . As before, $\underline{x}' \simeq \sum_{i=1}^I y'_i \phi'_i$. we compute the relation between the projection coefficients at the input pose v , and the projection coefficients at the reference pose r . We can then reconstruct the test face image at the reference pose (even though the test image is not in the training set). We assume a linear relation between the projection coefficients at the input pose v and the projection coefficients at the rotated or reference pose r as $y'_i = \underline{U} y_i$, where \underline{U} is a banded matrix whose non-zero elements are determined using a set of training images of individuals at poses v and r . Note that the linear MRDF transforms and the matrix \underline{U} are unique for a given input and reference pose.

It is quite trivial to generalize the above method to a limited possible set of V input poses. In each case

the relation between the input and reference pose is determined from training image pairs at the input and reference poses. After the system is trained, given an input unknown test face image at an unknown pose, we determine its pose using our discriminatory MRDFs (Sect. 2.1) and the modified k-NN classifier (Sect. 2.2) and then transform it to the reference pose. We chose the frontal face views as the reference pose in this work.

4 Face Recognition from Reconstructed Faces

The last stage of our pose-invariant face recognition procedure involves using the reconstructed faces at the reference pose for face recognition. Face recognition is a discrimination or classification process; hence we use the nonlinear discriminatory MRDF features similar to the one developed in Sect. 2.1. Given images of L individuals, we want to extract discriminatory features for L classes (each individual is a class). Therefore, the M best nonlinear features that will classify the L individuals are the M dominant eigenvectors of $[\sum_{i=1}^L \underline{C}_{i,r}]^{-1} [\sum_{i=1}^{L-1} \sum_{j=i+1}^L \underline{R}_{ij,r}]$, where $\underline{R}_{ij,r}$ is a higher-order correlation matrix that measures the separation between faces i and j at the reference pose, and $\underline{C}_{i,r}$ is the higher-order covariance matrix for face i at the reference pose. An efficient two-stage technique is used to calculate the features[3]. Note that this feature extraction procedure ensures that the output features of one individual are separated from the output features for another individual. This is quite useful if the reconstructed images for an individual from different input aspect views tend to be different from each other after pose transformation, and are also different from the actual image of the individual at the reference pose. For each training image, we use several different initial poses in designing our features; these different inputs will result in pose-transformed output images that differ slightly from each other, and from the actual face image at the reference pose. Our features are chosen to minimize these differences and to separate output features for different individuals.

5 Initial Results

We used the database provided by Foggio and Beymer (A.I. Lab, MIT) which consists of 62 individuals with 6 aspect views of each person (aspect view range with left/right movements upto $\pm 45^\circ$ (Fig. 2), one case with two degrees of freedom including a slight $\simeq 45^\circ$ upward motion (top, Fig. 2b). The head-on (reference) view is 0° . We used a leave-one-out technique to test our face pose estimation, pose transformation, and

face recognition results, since the face database is very small.

We tested our pose estimation procedure on all 6 sets of aspect views in the database. From the training set, we determined that a nonlinearity of $p=1.6$ provided the best discrimination between faces at all 6 poses using the criterion described in Sect. 2.1. We used $k=3$ in the modified k -NN classifier (Sect. 2) and 6 discriminatory nonlinear MRDF features. We obtained perfect (100%) pose estimation accuracy for all 62 face images at all 6 aspect views. These results show the effectiveness of the nonlinear MRDF for discrimination even when the number of classes (aspect views) is high and the faces are not in the training set.

We tested our pose transformation procedure (Sect. 3.2) using 3 input poses at $\pm 45^\circ$, $\pm 22.5^\circ$, and 45° with a $\pm 15^\circ$ upward movement. The head-on 0° view was the reference (rotated) view to which we transformed each input face. Details of the number of MRDFs used, etc. are provided in a previous publication [10].

The reconstructed faces at the 0° aspect views (bottom of Figs. 2a, b, c, and d) from different input aspect views (top Figs. 2a, b, c and d) look quite similar to the actual (unseen) faces at the 0° aspect view, even when facial hair is present (Figs. 2c and d) and when the test set contains other variations not present in the training set. Other methods have problems with such cases [3]. Further detailed descriptions of the facial pose transformation results are described in [10]. In a few cases, the pose-transformed faces are partly noisy, and blurred as in bottom of Figs. 2e and f, and hence differ from the corresponding actual face images at the head-on reference view. Therefore, the features that are used for face recognition should be robust to handle variations between the pose-transformed images of the same individual and the actual image.

To test classification performance, we used a nearest neighbor classifier and a leave one out technique. We calculated our nonlinear MRDF features for recognition/discrimination (Sect. 4) using all 6 aspect view images of 61 of the (training) faces. The mean (in MRDF feature space) of the pose-transformed versions of each of the 61 faces in the training set, and the 0° version of the 62nd face were used as the references/prototypes in the classifier. The poses of the 3 aspect views ($\pm 45^\circ$, $\pm 22.5^\circ$, and 45° with a $\pm 15^\circ$ upward movement) of the 62nd face (test set) were estimated, the faces were transformed to the reference 0° aspect view, the nonlinear MRDF features for recognition were extracted from each pose-transformed face, and each face was then classified using the nearest neighbor classifier in MRDF feature space.

This was repeated 62 times with a different individual left out each time. Perfect $P_C = 100\%$ recognition results were obtained using our nonlinear MRDF features as shown in Table 1 for all 3 input aspect views of each face. The results obtained using eigenfeatures (KL features) are also shown in Table 1 for purposes of comparison; their P_C is observed to be lower than those obtained using nonlinear MRDFs. These KL features were obtained by considering all $61 \times 6 + 1 = 367$ input pose-transformed training images as a single macro-class. Since the KL (PCA) features do not assign class-labels to each training sample, it maximizes the variance of the projections of all training sample images. Therefore, as seen in Table 1, no improvements in P_C are obtained even when a large number of KL features are used.

Acknowledgments

The authors gratefully acknowledge Thomas Poggio and David Beymer for providing the face database. The results are based upon nonlinear feature work supported by the Cooperative State Research, Education, and Extension Service, U.S. Department of Agriculture, under agreement No. 97-35503-4532, and active vision work supported by grant number No. IRI9530637 from the National Science Foundation.

References

- [1] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. PAMI*, 1(12):103-108, 1990.
- [2] R. Javidi and J. Li. Neural networks-based face recognition using Fourier plane nonlinear filters. In *Proc. SPIE*, volume 2365, pages 30-39, 1995.
- [3] R. Moghaddam, C. Nastar, and A. Pentland. Bayesian face recognition using deformable intensity surfaces. *MIT Media Lab Tech. Rept and Proc. CVPR 1996*, TR 371:1-7, 1996.
- [4] A. Pentland, R. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. *MIT Media Lab Tech. Rept and Proc. ICCV*, 1:84-91, 1994.
- [5] Ashit Talukder and David Casasent. Nonlinear features for pose estimation and classification of machined parts. In *Proc. SPIE: Intelligent Robots and Computer Vision XVII: Algorithms, Techniques, and Active Vision*, volume 3522, Nov. 1998.
- [6] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Trans. PAMI*, 19(7):733-742, Jul. 1997.

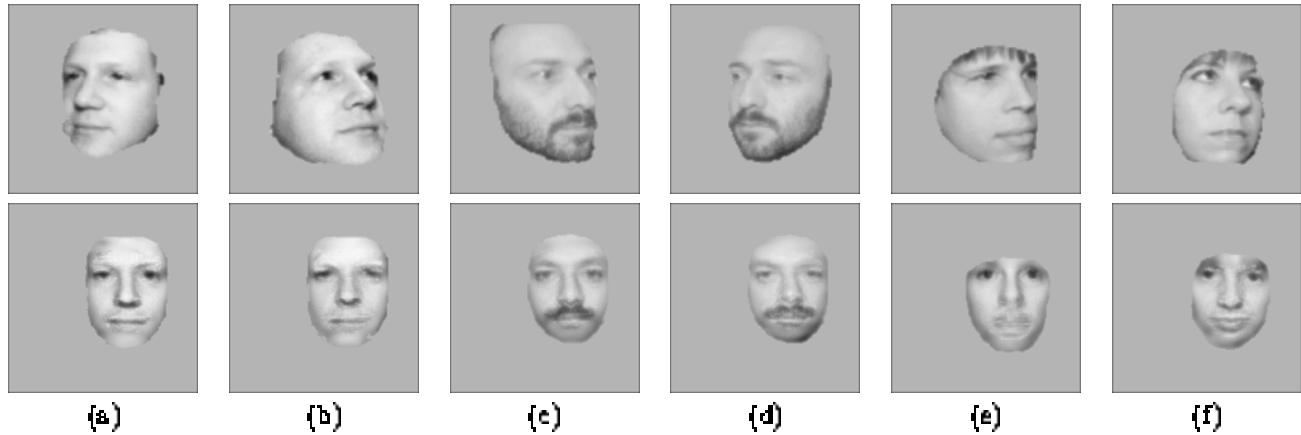


Figure 2: Face images at the input pose (top) and reconstructed at the reference pose (bottom).

- [7] A. Talukder and D. Casasent. General methodology for simultaneous representation and discrimination of multiple object classes. *Optical Engineering, Special Issue on Advanced Recognition Techniques*, 37(3):904–913, March 1998.
- [8] Aahit Talukder, David Casasent, Ha-Woon Lee, P. M. Keagy, and T. F. Schatzki. A new feature extraction method for classification of agricultural products from x-ray images. In *Proc. SPIE*, volume 3543, Nov. 1998.
- [9] T. Vetter and T. Poggio. Image synthesis from a single example image. In *Proceedings Computer Vision-ECCV*, volume 1, pages 652–659, May 1996.
- [10] Aahit Talukder and David Casasent. Pose estimation and transformation of faces from gray-scale images. In *Proc. SPIE: Intelligent Robots and Computer Vision XVII: Algorithms, Techniques, and Active Vision*, volume 3522, Nov. 1998.

Features	Input Test Pose				
	45°	22.5°	-22.5°	-45°	45° & up
	Recognition (P_C)%				
NL	100%	100%	100%	100%	100%
MRDF (8)					
Lin.	100%	100%	98.4%	100%	100%
MRDF (20)					
KL (8)	74.2%	85.5%	90.3%	74.2%	82.2%
KL (30)	87.1%	96.7%	96.7%	85.5%	90.3%
KL (100)	80.6%	83.9%	95.2%	80.6%	83.9%

Table 1: Face recognition results obtained on test images at different input aspect views pose-transformed to the reference aspect view.