3

Face Recognition

Marios Savvides, Jingu Heo, and Sung Won Park

Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
Marios.Savvides@ri.cmu.edu, jheo@andrew.cmu.edu, sungwonp@cmu.edu

3.1 Introduction

Robust face recognition systems are in great demand to help fight crime and terrorism. Other applications include providing user authentication for access control to physical and virtual spaces to ensure higher security. However the problem of identifying a person by taking an input face image and matching with the known face images in a database is still a very challenging problem. This is due to the variability of human faces under different operational scenario conditions such as illumination, rotations, expressions, camera view points, aging, makeup, and eyeglasses. Often, these various conditions greatly affect the performance of face recognition systems especially when the systems need to match against large scale databases. This low performance on face recognition prevents systems from being widely deployed in real applications (although many systems have been deployed, their use and accuracy is limited to particular operational scenarios) where errors like the false acceptance rate (FAR) and the false rejection rate (FRR) are considered in advance. FAR is the probability that the systems incorrectly accept an unauthorized person, while FRR is the probability that the systems wrongly reject an authorized person. In order to enhance the overall face recognition algorithm performance, numerous new algorithmic approaches such as Kernel Class-Dependent Feature Analysis (KCFA) [76] [32], Tensorfaces [71], manifold learning methods [58], kernel methods [66], and different Linear Discriminant Analysis (LDA) variants have been proposed [47] showing a great deal of improvement over conventional techniques. Among these, some of the new approaches such as KCFA, emphasize on the generalization to unseen people and Tensorfaces can deal with multiple factor analysis (different pose, illumination). Manifold learning methods can capture the underlying structures in the feature space of facial images. Traditional LDA variants try to find the best separation projection vectors by maximizing the Fisher's criteria [23].

Recently, 3D face recognition has gained attention in the face recognition community due to its inherent capability to overcome some of the traditional

problems of 2D imagery such as pose and lighting variation [16] [20] [24]. Commercial 3D acquisition devices can obtain a depth map (3D-shape) of the face. These usually require the user to be in very close proximity to the camera; additionally some devices will require the user to be still for several seconds for a good 3D model acquisition. In contrast 2D face acquisition can work from a distance and not require significant user co-operation. This is the trade-off with desiring to work with 3D shape data. One approach to address this issue is the 3D Morphable Model approach (3DMM) [14] which tries to recover a 3D face model from a 2D face image. 3DMM is an attractive method for handling different poses and illuminations effectively [12] [13] compared to other 3D face recognition approaches. However, this approach still requires significant ground-truth 3D face models for training the system and overall speed of rendering a 3D face takes several seconds. Other approaches to reconstruct the 3D-shape from 2D images of the face include Structure From Motion(SfM) [36]. A more comprehensive survey of recent 3D face recognition algorithms can be found at [63] [19] and the fusion of visual and thermal face recognition can be found at [38] reporting multi-model based face recognition systems lead to improved performance than single modality systems.

In this chapter we mainly focus on 2D-based face recognition approaches with some background on traditional methods in Section 2 and then emphasis on some of the newer approaches that tackle problems with 2D face recognition in the later sections. We briefly touch upon some of the most popularly used face databases by the face recognition community in Section 3. In the following sections, we describe the state-of-the-art techniques for face applications more deeply and show their experimental results. In Section 4, we show that KCFA can be successfully applied to face recognition with large scale challenging databases such as the Face Recognition Grand Challenge (FRGC) database [55] [1]. In section 5, we describe Tensorfaces for face recognition and novel face synthesis under different pose and illumination. Pre-processing approaches that aid 2D face recognition are dealt with in Section 6 detailing how to perform real-time pose correction using Active Appearance Models (AAMs). In Section 7, we show how to deal with very low-resolution, poor quality acquired face images with a novel super-resolution method that utilizes manifold learning techniques to achieve good reconstruction results. We then conclude with a short closing discussion in Section 8.

3.2 Face Recognition Techniques

Face recognition algorithms can be classified into two broad categories according to feature extraction schemes for face representation: feature-based methods and appearance-based methods [81]. Properties and geometric relations such as the areas, distances, and angles between the facial feature points are used as descriptors for face recognition. On the other hand, appearance-based methods consider the global properties of the face image intensity pattern.

Fig. 3.1. The first six basis vectors of Eigenfaces.

Typically appearance-based face recognition algorithms proceed by computing basis vectors to represent the face data efficiently. In the next step, the faces are projected onto these vectors and the projection coefficients can be used for representing the face images. Popular algorithms such as PCA, LDA, ICA, LFA, Correlation Filters, Manifolds and Tensorfaces are based on the appearance of the face. Holistic approaches to face recognition have trouble dealing with pose variations. Building image face mosaics like those in [65] [45] have been introduced to deal with the pose variation problem. We will also discuss in detail an Active Appearance Model approach to pose correction in later section. We review several of the popular face recognition algorithms as well as Elastic Bunch Graph Matching(EBGM) approach [75].

3.2.1 Eigenfaces (PCA)

Eigenfaces [69] also known as Principal Components Analysis (PCA) find the minimum mean squared error linear subspace that maps from the original N-dimensional data space into an M-dimensional feature space. By doing this, Eigenfaces (where typically $M << N$) achieve dimensionality reduction by using the M eigenvectors of the covariance matrix corresponding to the largest eigenvalues. The resulting basis vectors are obtained by finding the optimal basis vectors that maximize the total variance of the projected data(i.e. the set of basis vectors that best describe the data). The optimal basis PCA vectors $\mathbf{W}$ are the ones that maximize the following objective function

$$\mathbf{W}_{PCA} = \arg\max_{\mathbf{W}} |\mathbf{W}^T \mathbf{S}_T \mathbf{W}| = [w_1 \ w_2 \ \cdots \ w_m] \tag{3.1}$$

where $\mathbf{S}_T$ denotes the total scatter matrix which contains pixel-wise covariances of the face data. Figure 3.1 shows examples of Eigenfaces generated from the generic training images of FRGC dataset [1] after pre-processing the face images such as normalizing faces for rotation, scale and illumination compensation. PCA is good for data representation but not necessarily for class discrimination as we will discuss next.

3.2.2 Linear Discriminant Analysis (LDA) and Fisherfaces

Linear Discriminant Analysis (LDA) [23] is more suited for finding projections that best discriminate different classes. It does this by seeking the optimal projection vectors which maximize the ratio of the between-class scatter and the

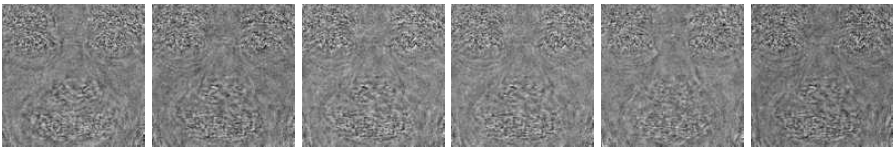

Fig. 3.2. The first six basis vectors of Fisherfaces.

within-class scatter (i.e. maximizing class separation in the projected space). The optimal basis vectors of LDA can be denoted as

$$\mathbf{W}_{LDA} = \arg\max_{\mathbf{W}} \frac{|\mathbf{W}^T \mathbf{S}_B \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_W \mathbf{W}|} \tag{3.2}$$

where $\mathbf{S}_B$ and $\mathbf{S}_W$ indicate between-class scatter matrix and within-class scatter matrix respectively.

Typically when dealing with face images (and most other image based pattern recognition problems) the number of training images is smaller than the number of pixels (or equivalently dimensionality of the data), thus the within-class scatter matrix $\mathbf{S}_W$ is singular causing problems for LDA [23]. To address this issue [10] first performs PCA to reduce the dimensionality of the data in order to overcome this singular-matrix problem and then applies LDA in this lower-dimensional PCA subspace. Improvement in recognition results was shown using this approach over traditional PCA. The projection vectors from Fisherfaces are those that maximize the following objective function:

$$\mathbf{W}_{Fisher} = \arg\max_{\mathbf{W}} \frac{|\mathbf{W}^T \mathbf{W}_{PCA}^T \mathbf{S}_B \mathbf{W}_{PCA} \mathbf{W}|}{|\mathbf{W}^T \mathbf{W}_{PCA}^T \mathbf{S}_W \mathbf{W}_{PCA} \mathbf{W}|} \tag{3.3}$$

Figure 3.2 shows examples of Fisherfaces generated from the generic training images of the FRGC dataset.

3.2.3 LDA variants

Direct LDA (DLDA) [78] derives eigenvectors using simultaneous diagonalization techniques. Unlike other LDA approaches [47], the DLDA simultaneously diagonalizes the between-class scatter matrix first and then diagonalizes the within-class scatter matrix. The eigenvectors with very small (close to zero) eigenvalues in the $\mathbf{S}_B$ can be discarded since they contain no discriminative power, while the eigenvectors with small eigenvalues of the $\mathbf{S}_W$ matrix simultaneously being kept, especially the null-space. Another LDA variant is called the Gram-Schmidt LDA (GSLDA) [82] approach avoids computing the inverse of the within-class scatter matrix or performing the diagonalization step needed in DLDA. These methods assert that the most discriminating power for LDA may lie in the null-space of the within scatter matrix which maximizes the Fisher's ratio.

3.2.4 Independent Component Analysis (ICA)

Independent Components Analysis (ICA) for face recognition has been applied in [8]. ICA seeks a non-orthogonal basis so that the transformed features are statistically independent, while PCA finds an orthogonal basis for face images so that the transformed features are uncorrelated. The basis images developed by PCA depend only on second-order statistics. ICA generalizes the concept of PCA to model higher-order statistical relationships. Original motivation for this decomposition comes from the need to separate audio streams into independent sources without prior knowledge of the mixing process.

3.2.5 Local Feature Analysis (LFA)

LFA [54] constructs a family of locally correlated feature detectors based on eigen-subspace decomposition. A selection or sparsification step produces a minimally correlated and topographically indexed subset of features that define the subspace of interest. Local representations offer robustness against variability due to changes in localized regions of the objects. The features used in the LFA method are less sensitive to illumination changes and are easier for estimating rotations. The LFA algorithm was used as a key component algorithm in FaceIt [56], which is one of the commercial face recognition systems.

3.2.6 Elastic Bunch Graph Matching (EBGM)

EBGM [75] constructs dynamic link architecture using image graphs to represent individual faces. An image graph representing a face image is a geometrical structure consisting of various nodes connected by edges. The nodes are located at facial landmarks such as the pupils and the corners of the mouth as shown in Figure 3.3. A set of training images is represented by the corresponding bunch of image graphs of those images. A set of complex Gabor wavelet coefficients (or Gabor jets) are used as local features at each node. These Gabor jets contain information of multiple orientations and frequencies for each node. When performing face recognition on a new facial image, each graph in the training set is matched to the image and the best match indicates the identity of person.

3.2.7 Neural Networks (NN) and Support Vector Machines (SVM)

Neural Networks and Support Vector Machines (SVMs) are usually used in low dimensional feature spaces due to the computational complexity of the processing involved using high-dimensional face data. Neural network approaches [43] have been widely explored for feature representation and face recognition. However, as the number of people for training increases, NN requires computational burden exponentially. Fusion of multiple neural networks

Fig. 3.3. The face model constructed by EBGM: (a) an image graph, (b) the facial landmarks of a test image detected by EBGM, (c) the image graph of a test image constructed by EBGM.

classifiers improved the overall performance of face recognition [27]. A face recognition system using hybrid neural and dual eigenspace methods has been proposed in [79]. However, in general it is not known what exactly the neural network has learned or how it will behave, and usually a significant amount of training data is required for good generalization which usually requires significant amount of offline training. Support Vector Machines (SVM) [70] [30] have been successfully applied for object recognition, by utilizing the kernel trick which maps data onto higher-dimensional feature spaces. The SVM finds the hyperplane that maximizes the margin of separation in order to minimize the risk of misclassification not only for the training samples, but to enable it to achieve better generalization to the unseen data.

3.2.8 Tensorfaces

Facial images have different appearance due to multiple factors such as variations across people, pose changes, lighting conditions and facial expressions. The Tensorfaces method [71] is proposed to model the variations of these factors by a multilinear framework. Tensors, which are higher-order extensions of matrices, allow us to construct multilinear models so as to analyze multiple factors of these facial variations. Lathauwer et al. [42] proposed Higher-Order Singular Value Decomposition (HOSVD) for tensor decomposition, which is an extension of Singular Value Decomposition (SVD) for matrix decomposition. Vasilescu et al. [71] introduced the idea of tensor decomposition into the area of computer vision and proposed Tensorfaces, a higher-order extension of the Eigenfaces method. By analyzing the tensor consisting of training images, the basis of each facial factor (expression, pose, etc.) in the training images can be obtained.

3.2.9 Manifolds

Learning the similarity among data points is one of the key concepts for the analysis of face images. In the previous work of face image analysis using

manifold learning methods, it has been shown that face images lie on a manifold [58] [28] [68] [11]. Also, it has been demonstrated that the variation of a certain facial factor such as various poses or expressions makes a sub-manifold in the manifold structure [29]. So, it is helpful to detect and analyze the underlying manifold structure in the distribution of facial image samples. Traditional methods such as PCA and LDA often see only the Euclidean structure, so they fail to discover the underlying structure if the data lies on a nonlinear manifold. The analysis of manifolds reveals the characteristics of the data distribution and can be applied for dimensionality reduction. Thus, to discover the nonlinear structure of manifolds, manifold learning techniques have been proposed [58] [68] [29]. In many real-world classification problems, the local manifold structure is more important than the global Euclidean structure. Thus, manifold learning techniques often use adjacency information among data samples to preserve the local manifold structure. By manifold learning techniques, neighboring points should still be in close proximity after mapping, and the points far from each other should still be far from each other in the new mapping.

3.2.10 Kernel Methods

Due to the large appearance changes in human face images, the linear subspace methods may not capture the non-linearity in facial image representation. As a result, the PCA and LDA algorithms have been extended to represent nonlinear mappings in a higher-dimensional space [9]. Computing and storing the new features in this higher-dimensional space becomes very expensive. Thus, the kernel trick is used for computational efficiency as it enables us to obtain the necessary inner products in the higher-dimensional feature space without computing the higher-dimensional feature mapping. Examples of kernel methods are Kernel Eigenfaces and Kernel Fisherfaces [77] . Kernel functions can be used without having to form an explicit high-dimensional mapping as long as kernels form an inner product space in this higher dimensional mapping and satisfy Mercer's theorem [51]. A number of papers combining linear subspace methods with the kernel trick including Kernel Direct LDA (KDLDA) [46], Kernel LDA (KDA) [50] or Kernel Fisher's Analysis (KFA), Kernel PCA (KPCA) [37], and Kernel ICA (KICA) [6] have been applied in face recognition showing improved performance over linear approaches.

3.2.11 Correlation Filters

Advanced correlation filter approaches such as those found in [40] [39] process images in the spatial frequency domain using closed form correlation filter solutions designed for specific optimization criteria. One of the most often used correlation filters is the Minimum Average Correlation Energy (MACE) [48] filter. This is designed to minimize the average correlation plane energy resulting from the training images, while constraining the correlation peak value

at the origin to pre-specified values. Correlation outputs from MACE filters typically exhibit sharp peaks, making the peak detection and location relatively easy and robust. Developed and applied originally in the field of Automatic Target Recognition (ATR), several different types of correlation filters have been proposed for face recognition in the presence of illumination variations [60], and occlusion [61], including a new hybrid shift-invariant PCA-correlation filter approach called Corefaces [62] that has all the subspace representation power of PCA and the shift-invariant and discrimination properties of advanced correlation filters.

3.3 Databases

There are several publicly available face databases for the research community to use for algorithm development, which provide a standard benchmark when reporting results. Different databases are collected to address a different type of challenge or variations such as illumination, pose, occlusion, etc. A more comprehensive review of available standard databases for developing face recognition algorithms can be found in the face database chapter of this book. In this section, several standard databases including PIE [64], FERET [2], FRGC [55], Yale [3] and AR [4] are briefly introduced to explain the experimental setup of the training and testing that lead to the different results reported in this chapter.

3.3.1 Face Recognition Grand Challenge (FRGC) database

The Face Recognition Grand Challenge (FRGC) conducted by the NIST is aimed at an objective and systematic evaluation of face recognition algorithms under different challenging conditions. Simultaneously, the aim for the FRGC is to push researchers to develop the next generation face recognition algorithms that can reduce the error rate in face recognition systems by an order of magnitude over the Face Recognition Vendor Test (FRVT) 2002 results [15] [56]. Details of the different FRGC experiments can be found at [55]. The FRGC data is partitioned into three datasets: a generic training set which one can use to train the face recognition system (if using PCA, this set is used to generate the PCA subspace), the target set (these are the images acquired under controlled conditions) and the probe set (the test set) captured under un-controlled conditions. The FRGC generic training set contains 12,776 images (from 222 subjects) taken under controlled and uncontrolled illuminations. The gallery set contains 16,028 images (from 466 subjects, with some overlap with subjects in the generic training set) under controlled illumination while the probe set contains 8,014 images (from 466 subjects) under uncontrolled illumination. The similarity matrix of matching scores between the target and probe sets are computed and reported to the NIST in the form of a $16,028 \times 8,014$ similarity matrix. Sample images of the FRGC database are shown in Figure 3.4.

(a) Controlled images (b) Uncontrolled images

Fig. 3.4. Sample images of the FRGC database.

(a) Sample poses among 13 poses

(b) Sample lighting conditions among 43 different lighting conditions

Fig. 3.5. Sample images of the PIE database.

3.3.2 FERET database

Prior to the FRGC, the NIST organized the FERET database and evaluation protocol [57] to facilitate the development of commercial face recognition systems. The FERET database is designed to measure the performance of face recognition algorithms on a large database in practical settings. The FERET program provides a large database of facial images taken from 1,199 individuals and collected between August 1993 and July 1996 to support algorithm development and evaluation. The FERET database consists of 14,126 images of 1,564 sets (1,199 original sets and 365 duplicate sets). For development purposes, 503 sets of images were released to the researchers, and the remaining sets were sequestered for independent evaluation.

3.3.3 Pose Illumination Expression (PIE) database

The CMU Pose, Illumination, and Expression (PIE) database [64] contains 41,368 facial images of 68 people. The images are acquired across different poses, under different illuminations, and with different facial expressions. First, in the CMU 3D Room, each person's images were captured under 13 different poses, 43 different illumination conditions, and 4 kinds of facial expressions. In particular, 43 different illumination conditions were obtained with only 21 flashes, since images were captured both with and without ambient background lighting switched on. Additionally each person has four types

of expressions: neutral expression, smiling, blinking, and talking. The CMU PIE database has been extensively used to analyze face images under different illumination and pose and for benchmarking the development of face recognition algorithms to handle such distortions.

3.3.4 AR database

The AR face database [4] was created by the Computer Vision Center (CVC), at Universitat of Autònoma de Barcelona. It contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women). The images acquired are frontal view pose with different facial expressions, illumination conditions, and occlusions (such as people wearing sun glasses and a scarf) making this database one of the more popular ones for testing face recognition algorithms in the presence of occlusion. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to the participants. Each person participated in two sessions, two weeks apart.

3.3.5 Yale Face database

The Yale database [3] contains 165 gray-scale images in GIF format of 15 individuals. There are 11 images per subject, one for each variation such as different facial expression, center-light, with glasses, happy, left-light, with and without glasses, normal, right-light, sad, sleepy, surprised, and wink. The Yale Face Database was extended to the Yale Face Database B, which contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses x 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured.

3.4 Advanced Correlation Filters

Due to their built-in shift invariance and designed distortion tolerance, advanced correlation filters are well suited for biometric verification/identification applications and have been shown to exhibit robustness to illumination variations and other distortions [60]. One of the popular filters called the Minimum Average Correlation Energy filter is designed to minimize the average correlation plane energy E resulting from the N training images defined as

$$E = \sum_{i=1}^{N} \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} c_i(x,y)^2 = \sum_{i=1}^{N} \sum_{u=0}^{M-1} \sum_{v=0}^{M-1} |C_i(u,v)|^2$$
$$= \sum_{i=1}^{N} \sum_{u=0}^{M-1} \sum_{v=0}^{M-1} |H(u,v)|^2 |X_i(u,v)|^2 = \mathbf{h}^+ \sum_{i=1}^{N} \mathbf{D}_i \mathbf{h} = \mathbf{h}^+ \mathbf{D} \mathbf{h}. \quad (3.4)$$

In Eq.(3.4), $c_i(x, y)$ is defined as the ith spatial correlation plane of size $M \times M$ which according to Parseval's theorem preserves the same energy in the Fourier frequency domain $C_i(u, v)$. $H(u, v)$ is the frequency domain filter and $X_i(u, v)$ is the 2D Fourier transform of the ith training image. $\mathbf{D}$ as defined in Eq.(3.4) is a $M^2 \times M^2$ diagonal matrix containing the average power spectrum of the training images along its diagonal. $^+$indicates the complex conjugate transpose. The MACE filter also specifies that correlation peak at the origin to pre-specified values represented below in the following equation:

$$\mathbf{X}^+\mathbf{h} = \mathbf{c} \tag{3.5}$$

where $\mathbf{X}$ is a $M^2 \times N$ complex valued matrix and its ith column contains the lexicographically re-ordered version of the 2D Fourier transform of the ith training image and $\mathbf{c}$ is a $N\times1$ row vector containing the correlation peaks desired for each of the N training images. Minimizing Eq.(3.4) the while satisfying the linear constraints in Eq.(3.5) yields a closed form solution to the optimization, giving the vectorized MACE filter $\mathbf{h}$ as

$$\mathbf{h} = \mathbf{D}^{-1}\mathbf{X}(\mathbf{X}^+\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{u}. \tag{3.6}$$

One of the recent advances in correlation filters is the class-dependent feature analysis (CFA) method which proposes a novel feature extraction method using correlation filters [33] [34] [5] [59]. Since the basis vectors acquired from either PCA or LDA are database dependent, it may be difficult to obtain basis vectors which represent or discriminate faces well on large databases. These approaches also exhibit poor generalization power; they may not discriminate faces well which have not been seen during training. Although kernel approaches such as KPCA, KLDA are attractive because of their ability to effectively use nonlinear mappings of face features, performance of these methods indicates room for improvement. Figure 3.6 shows a brief overview of our proposed work. Normalized face images are effectively mapped onto a high dimensional space using the kernel trick and features are extracted using correlation filters in the CFA framework, and then a kernel support vector machine (SVM) is designed among classes on this reduced feature set. We demonstrate our proposed work on the FRGC database to show the power of the CFA approach and how using this approach as a dimensionality reduction method further improves performance using SVM for classification in this feature space.

3.4.1 Kernel Class Dependent Analysis

In the CFA approach, one filter (e.g., MACE filter) is designed for each class in the generic training set. Then a test image $\mathbf{y}$ is characterized by the correlations of that test image with the n MACE filters, i.e.,

$$\mathbf{c} = \mathbf{H}^T\mathbf{y} = [\mathbf{h_{MACE-1}}\ \mathbf{h_{MACE-2}}\ \cdots\ \mathbf{h_{MACE-n}}]^T\mathbf{y} \tag{3.7}$$

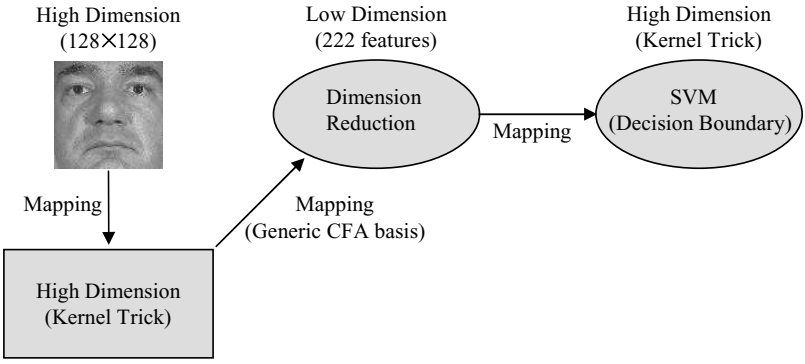

Fig. 3.6. An overview of the KCFA algorithm.

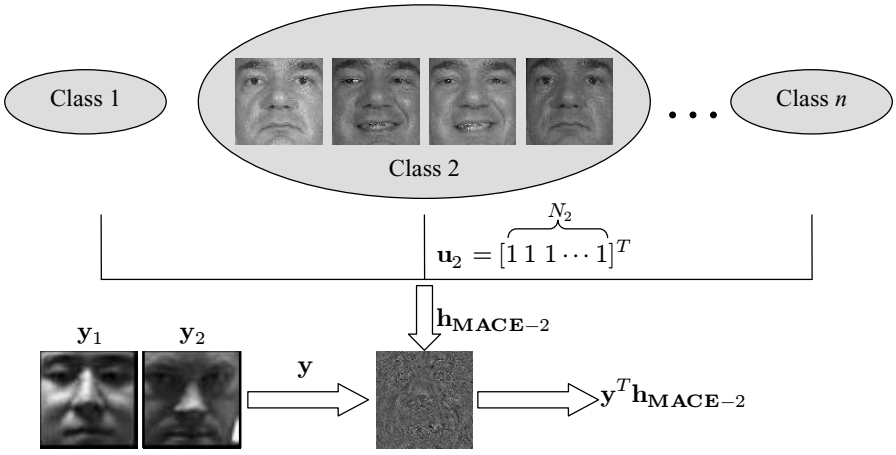

Fig. 3.7. The CFA algorithm; the filter response of $\mathbf{y}_1$ and $\mathbf{h}_{\mathbf{MACE}-2}$ can be distinctive to that of $\mathbf{y}_2$ and $\mathbf{h}_{\mathbf{MACE}-2}$.

where $\mathbf{h}_{\mathbf{MACE}-\mathbf{i}}$ is a filter designed for class i which is trained to give a small correlation output (close to 0) for all classes except for class i. For example, the number of filters generated by the FRGC generic training set is 222, since this generic training set contains 222 classes (or subjects). Then each input image $\mathbf{y}$ is projected onto those basis vectors to yield a 222 dimensional feature vector as shown in Figure 3.7 and N indicates the number of images per each class. The similarity of the probe image to the gallery image is computed in this 222 dimensional feature space.

Kernel functions defined by $\mathbf{K}(\mathbf{x}, \mathbf{y}) = \langle \mathbf{\Phi}(x), \mathbf{\Phi}(y) \rangle$ can be used without having to form the mapping explicitly, as long as the chosen kernel functions form an inner product space in this higher dimensional mapping and satisfy Mercer's theorem [70]. Examples of kernel functions are: Polynomial kernel ($\mathbf{K}(x, y) = (\langle \mathbf{x}, \mathbf{y} \rangle + 1)^p$), Radial Basis Function kernel ($\mathbf{K}(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} -$

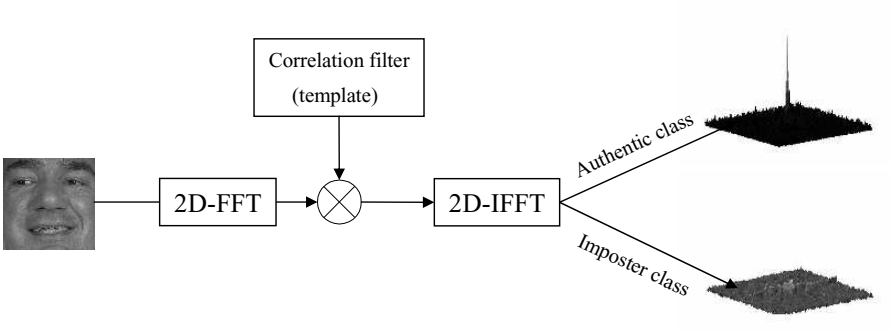

Fig. 3.8. The correlation peaks. The height of the peak indicates confidence of match, so we get a high peak for an authentic class at the origin with low correlation values in the remaining correlation plane. For imposter classes there is no discernible peak detected in the correlation plane.

$\mathbf{y}\|^2/2\sigma^2))$, and Neural-network kernel ($\mathbf{K}(\mathbf{x},\mathbf{y}) = \tanh(k\langle\mathbf{x},\mathbf{y}\rangle - \delta)$). The Kernel Correlation Filters can be extended from the linear correlation filters using the kernel trick. The kernel correlation output of a filter $\mathbf{h}$ and an input $\mathbf{y}$ can be expressed as

$$\begin{aligned}\mathbf{\Phi}(\mathbf{y}) \cdot \mathbf{\Phi}(\mathbf{h}) &= (\mathbf{\Phi}(\mathbf{y}) \cdot \mathbf{\Phi}(\mathbf{X}'))(\mathbf{\Phi}(\mathbf{X}') \cdot \mathbf{\Phi}(\mathbf{X}'))^{-1}\mathbf{u} \\ &= \mathbf{K}(\mathbf{y},\mathbf{x}'_\mathbf{i})\mathbf{K}(\mathbf{x}'_\mathbf{i},\mathbf{x}'_\mathbf{i})^{-1}\mathbf{u} \end{aligned} \tag{3.8}$$

where $\mathbf{X}'$ and $\mathbf{x}'_\mathbf{i}$ indicates the corresponding pre-processed versions of $\mathbf{X}$ and $\mathbf{x}_\mathbf{i}$. In the latest KCFA framework using different image representations and image resolutions combined using feature fusion, the pure 1-1 matching performance of KCFA has been improved to 82.4% verification at 0.1% FAR.

3.4.2 Support Vector Machines for Classification

A direct use of the SVM as a classifier on raw pixel data may not be practically feasible (for training) due to the large amount of available training data and large dimensionality of faces. Instead of using the SVM as a direct classifier on image pixels, we use SVMs in KCFA feature space. Since the dimensionality reduction based on KCFA is more efficient and discriminative than other approaches discussed, we use KCFA features (222 dimensional feature space) as an input for training the SVM. We design 466 SVMs (in a one-against all framework) using the gallery set of the FRGC data by changing the labeling information as shown in Figure 3.9. The probe images are then projected on the class-specific SVMs to provide a classification score between all the gallery images and all test images.

As shown in Figure 3.10, the distance measure using the SVM improves the results over the normalized cosine distance in the KCFA feature space and in many practical scenarios we can perform this as we will have access

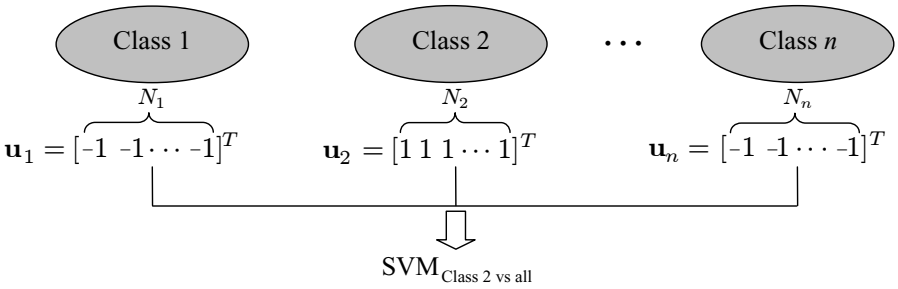

Fig. 3.9. Building decision boundary from class 2 vs. rest of all; **u** indicates the class labeling information vector and N depends on the number of images in each class.

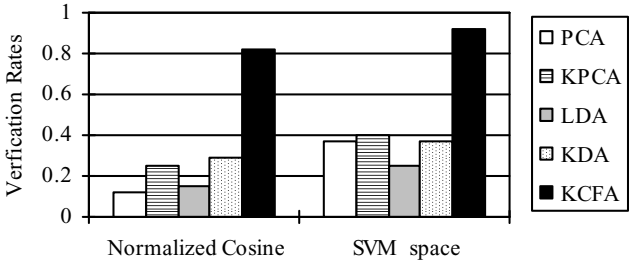

Fig. 3.10. Verification rate (VR) for the FRGC experiment 4 for different methods using the normalized cosine distance and the SVM space (0.1% FAR).

to the Target set. We also compared the different kernel approaches such as KPCA and KDA with different distance measures showing that the KCFA methods based on SVM have superior performance compared to other kernel approaches.

3.5 Tensorfaces

The Tensorfaces method is a novel method to analyze the facial appearance factors such as pose variations, lighting conditions and facial expressions. The traditional Tensorfaces method using multilinear algebra enables us to analyze these facial factors from a particular training set, but it has difficulties analyzing the factors of a new test image when all the factors of the test image are unknown or untrained. Thus, factorization methods for test images have been proposed [73] [52] [53].

3.5.1 Multilinear Analysis of Training Images

A tensor is a multilinear extension of a matrix; while a matrix deals with only two dimensions, a tensor can represent more than two dimensions. Higher-

Order Singular Value Decomposition (HOSVD) [42] is a higher-order extension of Singular Value Decomposition. HOSVD transforms an n-dimensional tensor into the weighted sum of the outer products of n independent vectors. When a training set has three facial factors such as people's identities, lighting conditions, and pose types, one of the general ways to analyze the training set is to construct a $N_{pixel} \times (N_{people} \times N_{light} \times N_{pose})$ matrix $\mathbf{D^{train}}$ which has the vectorized training images as columns. Here, N_{pixel} is defined as the number of pixels in one image, N_{people} is the number of people, N_{light} is the number of lighting conditions, and N_{pose} is the number of pose types in the training set. By using SVD, the matrix representing the training set is decomposed into two orthogonal bases and singular values:

$$\mathbf{D^{train}} = \mathbf{U_{pixel}} \mathbf{S} \mathbf{V}^T \tag{3.9}$$

where $\mathbf{U_{pixel}}$ is the basis of the column space and $\mathbf{V}$ is the basis of the row space of $\mathbf{D^{train}}$. By HOSVD, the same training set can be analyzed in more detail:

$$\mathbf{D^{train}} = \mathbf{U_{pixel}} \mathbf{Z} (\mathbf{U_{people}} \otimes \mathbf{U_{light}} \otimes \mathbf{U_{pose}})^T \tag{3.10}$$

where $\otimes$ represents the Kronecker product. This analysis of face images using HOSVD is called Tensorfaces [71]. The $N_{people} \times N_{people}$ matrix $\mathbf{U_{people}}$ spans the space of people parameters, the $N_{light} \times N_{light}$ matrix $\mathbf{U_{light}}$ spans the space of light parameters, and the $N_{pose} \times N_{pose}$ matrix $\mathbf{U_{pose}}$ spans the space of pose parameters. HOSVD can be represented by two forms; one is using tensors and tensor multiplications while the other is using matrices and the Kroneker products. Eq.(3.10) is the matrix notation of HOSVD. By the analogue of decomposing the training images, one image can be decomposed into the same kinds of factors, no matter whether it is in the training set or not:

$$\mathbf{d} = \mathbf{U_{pixel}} \mathbf{Z} (\mathbf{x_{people}} \otimes \mathbf{x_{light}} \otimes \mathbf{x_{pose}})^T. \tag{3.11}$$

In the case of the training image of the ith person, the jth lighting condition and the kth pose, the person's identity parameter $\mathbf{x_{people}}$ is the ith row of the matrix $\mathbf{U_{people}}$ since the $N_{people} \times 1$ vector $\mathbf{x_{people}}$ is the ith column of ${\mathbf{U_{people}}}^T$. For the same reason, the lighting parameter $\mathbf{x_{light}}$ of the training image is the jth row of the matrix $\mathbf{U_{light}}$, and $\mathbf{x_{pose}}$ is the kth row of the matrix $\mathbf{U_{pose}}$. The parameters of all the factors for the training image can be easily calculated by the multilinear analysis of the training set [72]. It is also easy to calculate the parameter $\mathbf{x_{people}}$ of a test image when it is the only unknown parameter and all the others are known [72] [74] or estimated by other techniques [67] [73]. However, it has difficulties obtaining the parameters $\mathbf{x_{people}}$, $\mathbf{x_{light}}$, and $\mathbf{x_{pose}}$ when all the parameters are unknown for a test image. In particular, if the test image has untrained pose or lighting conditions, it is more challenging to get the three parameters of the three factors. Consequently, the goal of factorization in the testing process is to solve for all

the unknown parameters $\mathbf{x_{people}}$, $\mathbf{x_{light}}$, and $\mathbf{x_{pose}}$ of any test image based on the multilinear analysis of the training set.

3.5.2 Multilinear Analysis of Testing Images

To factorize test images which have untrained poses or lighting conditions, a novel factorization method is proposed based on the previous work [52] [53]. By the method, all the factors can be estimated simultaneously without any *a priori* assumption or knowledge of the acquired face image. Moreover, the proposed method is applicable even when a test image has untrained lighting condition or pose. In the proposed method, to obtain factor parameters, first, it is shown that the tensor factorization problem can be formulated as a least squares problem with a quadratic equality constraint.

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{U_{pixel}}\mathbf{Z}\mathbf{x} - \mathbf{d^{test}}\|^2 \text{ subject to } \|\mathbf{x}\|^2 = 1 \tag{3.12}$$

where $\mathbf{d^{test}}$ is a given test image, and $\mathbf{x} = \mathbf{x_{people}} \otimes \mathbf{x_{light}} \otimes \mathbf{x_{pose}}$. The goal is to find $\hat{\mathbf{x}}$, the optimal value of $\mathbf{x}$ which minimizes the distance between the test image $\mathbf{d^{test}}$ and the reconstructed image by the estimated parameters.

Next, $\hat{\mathbf{x}}$ is obtained by well-defined numerical optimization techniques which allow us to obtain the facial appearance factors simultaneously. In [52], $\hat{\mathbf{x}}$ is estimated by the projection method [80], which is a better optimization scheme than Newton's method to solve the problem. After getting $\hat{\mathbf{x}}$, which is the Kronecker product of the mixing factors, it is decomposed into individual factors $\mathbf{x_{people}}$, $\mathbf{x_{light}}$, and $\mathbf{x_{pose}}$ by Higher-Order Power Method [41].

Table 3.1. The recognition rates using Tensor Factorization [52].

Methods	People and lightings conditions	People, lighting conditions and poses
EigenFaces	79.3%	69.4%
FisherFaces	89.2%	73.6%
Tensor factorization	95.6%	81.6%

Face recognition and synthesis are the main applications of tensor factorization, and results using the Yale Face Database B [3] can be found in [52]. The database contains 10 people, and each person has 65 different lighting conditions and 9 poses. Two kinds of multilinear models were constructed and tested; one is a bilinear model with two factors consisting of different people and lighting conditions, and the other is a trilinear model with three factors consisting of different people, lighting conditions, and poses. To select lighting variation for training, first, 10 lighting conditions among 65 were discarded since the images under the lighting conditions are so dark that it is

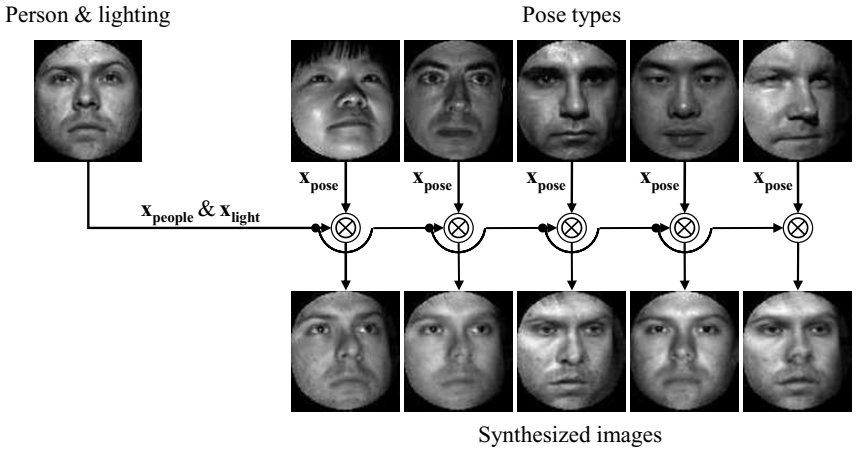

Fig. 3.11. The results of face synthesis on pose variation using tensor factorization with three factors: people, lighting conditions and poses. The synthesized images still have the person's identity parameter and the lighting parameter of the top left image, but they have different pose types.

hard to extract information from them. Next, every fifth sample was added to the training set. For the bilinear model, 11 lighting conditions of 10 subjects were used for training, while the other 44 lighting conditions were used for testing with no overlap between the two subsets. For the trilinear model, the lighting conditions were the same with the above bilinear model for both training and testing. Additionally, three poses are used for training while the other six poses are used for testing. Here, the three poses for training are the pose 0, 6, and 8 of the Yale Face Database; the pose 0 is the frontal pose, and the pose 6 and 8 are taken from the two largest degrees of the camera optical axis. Table 3.1 shows the results of face recognition, and Figure 3.11 shows the results of face synthesis on pose variation.

3.6 Active Appearance Models for Face Recognition

An Active Appearance Model (AAM) [21] [22] is a statistical model to interpret (in this case) facial images with known parameters. It is comprised of a shape model and an appearance (pixel intensity) model using PCA. The general procedure of an AAM algorithm can be explained as follows. The manually labeled (during training only) shape information (2D vertices) is normalized to deal with global geometric transformations such as scale and rotation using Procrustes Analysis which is a well-known technique for analyzing the statistical distribution of shapes. Then, the normalized shape ($\mathbf{s}$) for an AAM model can be expressed by the mean shape ($\bar{\mathbf{s}}$) and a linear combination of the basis vectors with the amount of projection coefficient vector($\mathbf{p_s}$)

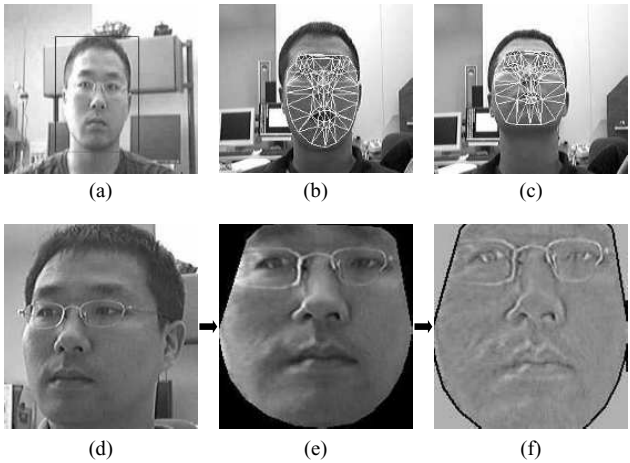

Fig. 3.12. Overview of the integrated face recognition system from video: (a) face detection, (b) AAM fitting, (c) automatic AAM tracking of fiducial facial landmarks, (d) a non-frontal face pose image, (e) face pose correction normalization, and (f) an illumination removal processed image (notice the removal of left cheek specular reflections).

which can be denoted by:

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{V_s p_s} \tag{3.13}$$

where $\mathbf{V_s}$ indicates the eigenvector matrix of the shapes. After warping an original image based on the mean shape, the appearance (shape free) model also can be represented by the mean appearance ($\bar{A}(\bar{\mathbf{s}})$) and a linear combination of appearance basis vectors. This can be expressed by:

$$A(\bar{\mathbf{s}}) = \bar{A}(\bar{\mathbf{s}}) + \mathbf{V}_A \mathbf{p}_A \tag{3.14}$$

where $A(\bar{\mathbf{s}})$ indicates a vectorized appearance image after warping based on the mean shape, $\mathbf{V}_A$ indicates the eigenvector matrix of the appearances and $\mathbf{p}_A$ is the projection coefficient vector. For fitting an AAM into new images, we need to minimize the distance between new images with known model parameters. Then, the objective function can be denoted as

$$E(\mathbf{s}) = \|A(\bar{\mathbf{s}}) - I(\mathbf{s}, \bar{\mathbf{s}})\|^2 \tag{3.15}$$

where $I(\mathbf{s}, \bar{\mathbf{s}})$ indicates an input image with shape s which is being warped based on the mean shape ($\bar{\mathbf{s}}$). The minimization can be done by assuming a linear relationship between residual errors (δI) and displacement vectors (δc) [21].

By using the current residual errors, the iterative model refinement procedure is applied to find the direction which gives the minimum residual error. To speed this fitting process, the inverse composition algorithm is proposed

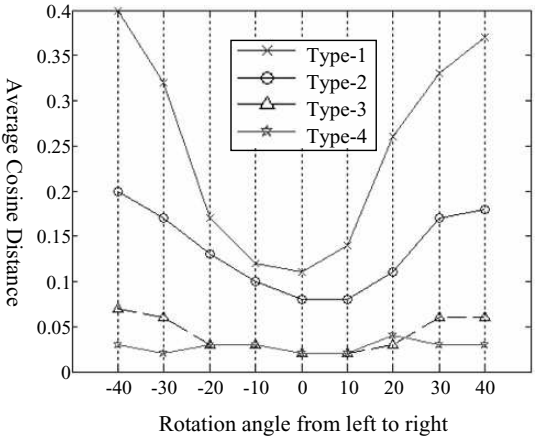

Fig. 3.13. Face matching distance performance across poses.

Table 3.2. Database for performance evaluation

Database (Gallery(G))	Test (Query(Q))
1 frontal image, 10 people, (10 images)	30,40 images, 10 people (376 images)

in [49] [35] [26]. They use the inverse compositional warp update rather than updating $c \leftarrow c + \delta c$, which leads to fast fitting convergence and low computation cost. Typically, for the video-based AAM, the first frame is selected for the shape approximation. Then the estimated shape on the first frame can be the initialization of the shape on the next frame assuming the shape does not change dramatically from frame to frame (high frame rate video capture (>7 fps)). If the first video frame fails to estimate the correct face shape on the image, then the next consecutive frame will also highly likely fail to converge to estimate the correct face shape on the image. In the initialization step in searching for these points, we automatically run a face detector [63] as shown in Figure 3.12 (a) and (b) for providing a good initial region-of-interest search space. Once the AAM fits on the detected face fiducial points and starts tracking, the faces are then warped into neutral frontal pose faces. Finally they are then passed through an illumination-pre-processing step before entering a face recognition matcher system as shown in Figure 12 (e) and (f). By converting non-frontal images into frontal pose and applying illumination compensation, the intra-variations of the same individuals are greatly reduced. Therefore more suitable for inputting these to a face recognition system for robust matching under different pose and illumination conditions [31].

Table 3.2 shows the database for evaluating the face matching performance. From a real video sequence, we captured about 30 images of variyng pose angles per person and measured the matching distance under different preprocessing techniques shown in the paper. Figure 3.13 shows the aver-

(a) Original images with pose changes

(b) Pose correction into frontal images using symmetry

Fig. 3.14. The results of pose, expression, and illumination normalization using AAM. Each column image in (b) results from the image of the same column in (a).

age matching similarity considering only the pose changes from left to right with different scenario types. Four different types of experiments are evaluated using original raw pose images denoted here (Type-1), AAM pose corrected warped images denoted as (Type-2), and illumination processed images (Type-3), and symmetry imposed images (Type-4) for large pose angles. By converting non-frontal faces into frontal face without expressions, the average distance between same individuals is greatly reduced. These results show that any face recognition algorithm can benefit by applying these approaches as a face-image pre-processing step. More examples of converting non-frontal images into frontal images are shown in Figure 3.14.

3.7 Face Super-resolution using Locality Preserving Projections

Recent work has shown that face images lie on a manifold [58] [28] [68] [11], thus we expected that manifold learning methods in general can improve the analysis of facial images and applications such as face recognition, super-resolution and face synthesis. Based on this idea, Chang et al. [18] developed the Neighbor Embedding algorithm for face super-resolution. They assume that the local distribution structure in sample space is preserved in the down-sampling process, and apply one of the manifold learning methods, Locally Linear Embedding (LLE) [58]. However, most of the manifold learning methods such as LLE have difficulty generating mapping functions for new test images. Thus, manifold analysis so far faced some limitations when applied to face super-resolution problem. Moreover, the objective of super-resolution

is to recover a high-dimensional image from a low-dimensional one, where as manifold learning methods are more suited for the opposite,namely dimensionality reduction. LPP tries to find a linear projective mapping for dimensionality reduction. Compared to LPP, other manifold learning techniques such as Isomap Tenenbaum et al.(2003), LLE [58], or Laplacian Eigenmap [11] define the mapping only on the training data. They successfully show the training data are distributed along manifolds, but it is unclear how to evaluate these mappings for new test samples. On the other hand, by using LPP, we can obtain a well-defined transformation matrix which can be applied to a new set of test images which are absent from the training set. In this framework, we apply LPP for every image face patch to model a high-resolution patch $\mathbf{x}_H$ from a down-sampled face patch $\mathbf{x}_L$. Given a high-resolution patch, the corresponding low-resolution patch is computed by down-sampling:

$$\mathbf{x}_L = \mathbf{B}\mathbf{x}_H \tag{3.16}$$

where $\mathbf{B}$ is the transformation matrix for the mapping from high-resolution to low-resolution. LPP aims to find a low dimensional embedding from a high dimensional patch, so it can be used for dimensionality reduction. LPP performs dimensionality reduction by projecting a high dimensional vector onto a low dimensional subspace. However, LPP can be also applied to the super-resolution problem; which is to map a low-resolution patch onto a high-resolution patch subspace. In this case, we must be able to obtain the coefficients in the high-resolution space from a given low-resolution patch. To do this estimation, various probabilistic approaches such as Belief Propagation have been employed to infer the coefficients of a high-resolution patch from a low-resolution patch [44] [17] [25]. In our case, we employ a Maximum *a posterior* (MAP) estimator to find the LPP coefficients of a high-resolution patch. A MAP approach has also been used to estimate the PCA coefficients of a high-resolution patch from a low-resolution one [44] [25]. Given the patches taken from training images, the LPP coefficients $\mathbf{y}_H$ are calculated by

$$\mathbf{y}_H = \mathbf{A}^T\mathbf{x}_H, \mathbf{x}_H = \mathbf{A}\mathbf{y}_H \tag{3.17}$$

where $\mathbf{A}$ is the projective matrix of LPP in Eq.(3.17). Maximizing $p(\mathbf{x}_L|\mathbf{x}_H)$ $p(\mathbf{x}_H)$ in Eq.(3.17) is equivalent to maximizing $p(\mathbf{x}_L|\mathbf{y}_H)p(\mathbf{y}_H)$. The prior $p(\mathbf{y}_H)$ is modeled by Gaussian distribution function:

$$p(\mathbf{y}_H) = \frac{1}{Z}\exp\left(-\mathbf{y}_H^T\mathbf{\Lambda}^{-1}\mathbf{y}_H\right) \tag{3.18}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2)$ and Z is a normalization constant. The likelihood $p(\mathbf{x}_L|\mathbf{x}_H)$ is denoted by

$$p(\mathbf{x}_L|\mathbf{x}_H) = \frac{1}{Z}\exp\left(-\|\mathbf{B}\mathbf{A}\mathbf{y}_H - \mathbf{x}_L\|^2/\lambda\right). \tag{3.19}$$

To maximize $p(\mathbf{x}_L|\mathbf{x}_H)p(\mathbf{x}_H)$, the optimal $\mathbf{y}_H$ is given by

(a) Input 24×32 (b) Original 96×128 (c) Cubic-B spline

(d) Baker et al. (e) Our method

Fig. 3.15. The results of face super-resolution using LPP and other methods.

$$\mathbf{y}_H^* = (\mathbf{A}^T\mathbf{B}^T\mathbf{B}\mathbf{A} + \lambda\mathbf{\Lambda}^{-1})^{-1}\mathbf{A}^T\mathbf{B}^T\mathbf{x}_L \tag{3.20}$$

where λ is decided empirically. If λ is too small, $\mathbf{x}_H$ cannot be obtained because $\mathbf{A}^T\mathbf{B}^T\mathbf{B}\mathbf{A}$ is close to singular.

Figure 3.15 shows comparative results using our method and the other competing approaches [7]. Note that our approach results in a more detailed reconstruction, producing a better psychovisual approximation to the original image compared to the other approaches.

3.8 Conclusion

Robust face recognition systems should be able to handle the variations that occur under practical operational scenarios. This means having the ability to handle any and all of face variations under different lighting, pose, expressions, and other variation factors such as low resolution face acquisition from a distance. To improve the performance and address each variation, numerous new algorithms have been proposed aiming at generalization to unseen people, multiple factor analysis and hidden structures of the faces. In case of low resolution faces, pre-processing methods that can enhance the resolution of face images have been detailed. Small pose variations can be handled and trained by different classifiers; however large pose variations can only be modelled by methods such as Tensorfaces and our proposed extensions. Furthermore

we show that using computer vision approaches such as Active Appearance Models (AAMs) can be used to model, track facial features to warp back a frontal facial pose as input to traditional classifiers. We have shown that using this and a multi-view AAM approach, we can handle even nearly completely profile views if we assume facial symmetry.

At the core classifier level, many algorithms exist for researchers to choose from. The FRGC database is currently the largest database with the most challenging practical variations which challenge simple algorithms such as PCA by yielding only 12% verification at 0.1% FAR. The new KCFA algorithm approach is promising, performing very efficient dimensionality reduction. Performance has been demonstrated using 222 features only and achieved a maximum verification rate of 82.4% at 0.1% FAR. Having such a low dimensionality allows for faster database searching, with ability to search among millions in a matter of seconds or less. Furthermore, this data is very compact that can easily fit on a smart-card or e-passport for verification.

3.9 Acknowledgments

This research is funded in part by the United States Technical Support Working Group (TSWG) and in part by Carnegie Mellon CyLab.

References

1. `http://www.frvt.org/FRVT2006/default.aspx`.
2. `http://www.itl.nist.gov/iad/humanid/colorferet/home.html`.
3. `http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html`.
4. `http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html`.
5. R. Abiantun, M. Savvides, and B. V. K Vijayakumar. How Low Can You Go?Low Resolution Face Recognition Study Using Kernel Correlation Feature Analysis on FRGC dataset. In *Proceedings of the IEEE Biometrics Symposium*, September 2006.
6. F. R. Bach and M. I. Jordan. Kernel Independent Component Analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.
7. S. Baker and T. Kanade. Hallucinating Faces. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 83–88, 2000.
8. M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, 2002.
9. G. Baudat and F. Anouar. Generalized Discriminant Analysis Using a Kernel Approach. *Neural Computation*, 12(10):2385–2404, 2000.
10. P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.
11. M. Belkin and P. Niyogi. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. *Advances in Neural Information Precessing Systems 14, MIT Press, Cambridge*, pages 585–591, 2002.

12. V. Blanz, S. Romdhani, and T. Vetter. Face Identification across Different Poses and Illuminations with a 3D Morphable Model. In *Proceedings of Fifth Int'l Conference Automatic Face and Gesture Recognition*, pages 202–207, 2002.
13. V. Blanz and T. Vetter. A Morphable Model for the Synthesis of 3D Faces. In *Computer Graphics Proceedings of SIGGRAPH*, pages 187–194, 1999.
14. V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.
15. M. Bone and D. Blackburn. Face Recognition at a Chokepoint: Scenario Evaluation Results. Evaluation report, Department of Defense, 2002.
16. K. W. Bowyer, K. Chang, and P. Flynn. An Evaluation of Multimodal 2D+3D Face Biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, April 2005.
17. D. Capel and A. Zisserman. Super-resolution from multiple views using learnt image models. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 627–634, 2001.
18. H. Chang, D. Y. Yeung, and Y. Xiong. Super-resolution through Neighbor Embedding. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 275–282, 2004.
19. K. Chang, K. Bowyer, and P. Flynn. A Survey of Approaches and Challenges in 3D and Multi-Modal 2D+3D Face Recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006.
20. X. Chen, P. Flynn, and K. Bowyer. IR and Visible Light Face Recognition. *Computer Vision and Image Understanding*, 99(3):332–358, September 2005.
21. T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *Proceedings of the European Conference on Computer Vision*, volume 2, pages 484–498, 1998.
22. T. F. Cootes, K. Walker, and C. J. Taylor. View-Based Active Appearance Models. In *Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition*, pages 227–232, March 2000.
23. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification.* Wiley-Interscience Publication, 2000.
24. P. Flynn and A. Jain. BONSAI: 3D Object Recognition Using Constrained Search. *Computer Vision and Image Understanding*, 13(10):1066–1075, October 1991.
25. W. T. Freeman and E. C. Pasztor. Learning low-level vision. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1182–1189, 1999.
26. R. Gross, S. Baker, I. Matthews, and T. Kanade. Face Recognition Across Pose and Illumination. In Stan Z. Li and Anil K. Jain, editors, *Handbook of Face Recognition.* Springer, New York, 2004.
27. S. Gutta, J. R. J. Huang, P. Jonathon, and H. Wechsler. Mixture of experts for classification of gender, ethnic origin, and pose of human faces. *IEEE Transactions on Neural Networks*, 11(4):948–960, 2000.
28. X. He and O. Niyogi. Locality preserving projections. *Advances in Neural Information Processing Systems*, 16, 2003.
29. X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face Recognition Using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3), March 2005.

30. B. Heisele, P. Ho, and T. Poggio. Face Recognition with Support Vector Machines: Global versus Component-based Approach. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pages 688–694, 2001.
31. J. Heo and M. Savvides. Real-time Face Tracking and Pose Correction for Face Recognition using Active Appearance Models. *The International Society of Optical Engineering(SPIE) (to appear)*, 2007.
32. J. Heo, M. Savvides, R. Abiantun, C. Xie, and B. V. K. Vijaya Kumar. Face Recognition with Kernel Correlation Filters on a Large Scale Database. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2006.
33. J. Heo, M. Savvides, and B. V. K. Vijaya Kumar. Large scale face recognition with distance measure in support vector machine spaces. In *Proceedings of the International conference on Enterprise Information systems*, 2006.
34. J. Heo, M. Savvides, C. Xie, and B. V. K. Vijaya Kumar. Which Facial Regions are Discriminative? Partial Face Recognition of FRGC data using Support Vector Machine Kernel Correlation Feature Analysis. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.
35. C. Hu, J. Xiao, I. Matthews, S. Baker, J. Cohn, and T. Kanade. Fitting a Single Active Appearance Model Simultaneously to Multiple Images. In *Proceedings of the British Machine Vision Conference*, September 2004.
36. T. Kanade and D. D. Morris. Factorization Methods for Structure From Motion. *Philosophical Transactions of the Royal Society of London, Series A*, 356(1740):1153–1173, 1998.
37. K. I. Kim, K. Jung, and H. J. Kim. Face recognition using kernel principal component analysis. *IEEE Signal Processing Letters*, 9(2):40–42, 2002.
38. S. G. Kong, J. Heo, F. Boughorbel, Y. Zheng, B. R. Abidi, A. Koschan, M. Yi, and M. A. Abidi. Adaptive Fusion of Visual and Thermal IR Images for Illumination-Invariant Face Recognition. *International Journal of Computer Vision, Special Issue on Object Tracking and Classification Beyond the Visible Spectrum*, 71(2):215–233, February 2007.
39. B. V. K. Vijaya Kumar. Tutorial survey of composite filter designs for optical correlators. *Applied Optics*, 31(2):4773–4801, 1992.
40. B. V. K. Vijaya Kumar, Abhijit Mahalanobis, and Richard D. Juday. *Correlation Pattern Recognition*. Cambridge University Press, UK, November 2005.
41. L. D. Lathauwer, P. Common, B. D. Moor, and J. Vandewalle. Higher-order power method - application in independent component analysis. In *Proceedings of the International Symposium on Nonlinear Theory and its Applications (NOLTA'95)*, pages 91–96, 1995.
42. L. D. Lathauwer, B. D. Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal of Matrix Analysis and Applications*, 21(4):1253–1278, 2000.
43. S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. Face Recognition: A Convolutional Neural-Network Approach. *IEEE Transactions on Neural Networks*, 8(1):98–113, 1997.
44. C. Liu, H. Shum, and C. Zhang. A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 192–198, 2001.

45. X. Liu and T. Chen. Pose Robust Face Recognition Based on Mosaicing –An Example Usage of Face In Action (FIA) Database. In *Demo session of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2004.
46. J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithms. *IEEE Transactions on Neural Networks*, 14(1):117–126, 2003.
47. J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Face recognition using LDA-based algorithms. *IEEE Transactions on Neural Networks*, 14(1):195–200, 2003.
48. A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent. Minimum average correlation energy filters. *Applied Optics*, 26:3633–3630, 1987.
49. I. Matthews and S. Baker. Active Appearance Models Revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.
50. S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Muller. Fisher Discriminant Analysis with Kernels. In *Proceedings of IEEE Neural Networks for Signal Processing Workshop*, 1999.
51. K. R. Muller, K. Tsuda S. Mika, G. Ratsch, and B. Scholkopf. Introduction to Kernel-Based Learning Algorithms. *IEEE Transactions on Neural Network*, 12(2):181–202, 2001.
52. S. W. Park and M. Savvides. Estimating mixing factors simultaneously in multilinear tensor decomposition for robust face recognition and synthesis. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop*, 2006.
53. S. W. Park and M. Savvides. Individual TensorFace Subspaces For Efficient and Robust Face Recognition That Do Not Require Factorization. In *Proceedings of the Biometrics Symposium 2006*, September 2006.
54. P. S. Penev and J. J. Atick. Local Feature Analysis: A General Statistical Theory for Object Representation. *Network: Computation in Neural Systems*, 7(3):477–500, 1996.
55. P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Fisher Discriminant Analysis with Kernels. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 947–954, 2005.
56. P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone. Face Recognition Vendor Test 2002. Evaluation report, National Institute of Standards and Technology, 2003.
57. P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
58. S. Roweis and L. K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, 2000.
59. M. Savvides, J. Heo, R. Abiantun, C. Xie, and B. V. K. Vijaya Kumar. Class Dependent Kernel Discrete Cosine Transform Features for Enhanced Holistic Face Recognition in FRGC-II. In *IEEE Conference on Acoustic, Speech, and Signal Processing*, 2006.
60. M. Savvides, B. V. K. Vijaya Kumar, and P. Khosla. Face verification using correlation filters. In *Proceedings of Third IEEE Automatic Identification Advanced Technologies*, pages 56–61, Tarrytown, NY, 2002.

61. M. Savvides, B. V. K. Vijaya Kumar, and P. Khosla. Robust, Shift-Invariant Biometric Identification from Partial Face Images. In *Biometric Technologies for Human Identification, SPIE Defense and Security Symposium*, volume 5404, pages 124–135, 2004.
62. M. Savvides, B. V. K. Vijaya Kumar, and P. Khosla. Corefaces: A Shift-Invariant Principal Component Analysis (PCA) Correlation Filter bank for Illumination-Tolerant Face Recognition. *Face Biometrics for Personal Identification Multi-Sensory Multi-Modal Systems*, 2007.
63. A. Scheenstra, A. Ruifrok, and R. C. Veltkamp. A Survey of 3D Face Recognition Methods. *Lecture Notes in Computer Science, Springer*, 3546:891–899, 2005.
64. T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination, and Expression (PIE) Database. In *Proceedings Fifth Intl Conference Automatic Face and Gesture Recognition*, 2002.
65. R. Singh, M. Vatsa, A. Ross, and A. Noore. Performance Enhancement of 2D Face Recognition via Mosaicing. In *Proceedings of 4th IEEE Workshop on Automatic Identification Advanced Technologies (AutoID)*, pages 63–68, 2005.
66. J. S. Taylor and N. Cristianni. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
67. J. B. Tenenbaum and W. T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12:1246–1283, 2000.
68. J. B. Tenenbaum, V. Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(12):2319–2323, 2003.
69. M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
70. V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.
71. M. A. O. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: TensorFaces. In *Proceedings European Conference on Computer Vision*, pages 447–460, 2002.
72. M. A. O. Vasilescu and D. Terzopoulos. Multilinear subspace analysis of image ensembles. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 93–99, June 2002.
73. M. A. O. Vasilescu and D. Terzopoulos. Multilinear independent components analysis. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 547–553, June 2005.
74. H. Wang and N. Ahuja. Facial expression decomposition. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pages 958–965, 2003.
75. L. Wiskott, J. M. Fellous, N. Krüger, and C. von der Malsburg. Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.
76. C. Xie, M. Savvides, and B. V. K. Vijaya Kumar. Kernel Correlation Filter Based Redundant Class-Dependence Feature Analysis (KCFA) on FRGC2.0 Data. In *IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, pages 32–43, 2005.
77. M. H. Yang. Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition using Kernel Methods. In *IEEE Conference on Automatic Face and Gesture Recognition*, pages 215–220, 2002.
78. H. Yu and J. Yang. A direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recognition*, 34(10):2067–2070, 2001.

79. D. Zhang, H. Peng, J. Zhou, and S. K. Pal. A novel face recognition system using hybrid neural and dual eigenspaces methods. *IEEE Transactions on Systems, Man and Cybernetics - Part A*, 32(6):787–793, 2002.
80. Z. Zhang and Y. Huang. A Projection Method for Least Squares Problems with a Quadratic Equality Constraint. *Journal of Matrix Analysis and Applications*, 25(1):188–212, 2003.
81. W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face Recognition: A Literature Survey. *ACM Computing Surveys*, pages 399–458, 2003.
82. W. Zheng, C. Zou, and L. Zhao. Real-Time Face Recognition Using Gram-Schmidt Orthgonalization for LDA. In *IEEE Conference of Pattern Recognition*, pages 403–406, 2004.