# Gabor-based Kernel PCA with Fractional Power Polynomial Models for Face Recognition

Chengjun Liu *

**Abstract**— This paper presents a novel Gabor-based kernel Principal Component Analysis (PCA) method by integrating the Gabor wavelet representation of face images and the kernel PCA method for face recognition. Gabor wavelets first derive desirable facial features characterized by spatial frequency, spatial locality, and orientation selectivity to cope with the variations due to illumination and facial expression changes. The kernel PCA method is then extended to include fractional power polynomial models for enhanced face recognition performance. A fractional power polynomial, however, does not necessarily define a kernel function, as it might not define a positive semi-definite Gram matrix. Note that the sigmoid kernels, one of the three classes of widely used kernel functions (polynomial kernels, Gaussian kernels, and sigmoid kernels), do not actually define a positive semi-definite Gram matrix, either. Nevertheless, the sigmoid kernels have been successfully used in practice, such as in building support vector machines. In order to derive real kernel PCA features, we apply only those kernel PCA eigenvectors that are associated with positive eigenvalues. The feasibility of the Gabor-based kernel PCA method with fractional power polynomial models has been successfully tested on both frontal and pose-angled face recognition, using two data sets from the FERET database and the CMU PIE database, respectively. The FERET data set contains 600 frontal face images of 200 subjects, while the PIE data set consists of 680 images across 5 poses (left and right profiles, left and right half profiles, and frontal view) with 2 different facial expressions (neutral and smiling) of 68 subjects. The effectiveness of the Gabor-based kernel PCA method with fractional power polynomial models is shown in terms of both

*Chengjun Liu is with the Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102. E-mail: *liu@cs.njit.edu*.

absolute performance indices and comparative performance against the PCA method, the kernel PCA method with polynomial kernels, the kernel PCA method with fractional power polynomial models, the Gabor wavelet based PCA method, and the Gabor wavelet based kernel PCA method with polynomial kernels.

**Index Terms** — Face recognition, fractional power polynomial models, Gabor wavelet representation, Gabor-based kernel PCA method, kernel Principal Component Analysis (PCA)

# 1 Introduction

Face recognition involves computer recognition of personal identity based on geometric or statistical features derived from face images [11], [8], [48], [6], [40]. Even though humans can detect and identify faces in a scene with little or no effort, building an automated system that accomplishes such objectives is very challenging. The challenges are even more profound when one considers the large variations in the visual stimulus due to illumination conditions, viewing directions or poses, facial expression, aging, and disguises such as facial hair, glasses or cosmetics. Face recognition research provides the cutting edge technologies in commercial, law enforcement, and military applications. An automated vision system that performs the functions of face detection, verification and recognition will find countless unobtrusive applications, such as airport security and access control, building (embassy) surveillance and monitoring, human-computer intelligent interaction and perceptual interfaces, smart environments at home, office, and cars [35], [36], [11], [8], [40].

This paper presents a novel Gabor-based kernel Principal Component Analysis (PCA) method by integrating the Gabor wavelet representation of face images and the kernel PCA method for face recognition. Gabor wavelets [31], [13] first derive desirable facial features characterized by spatial frequency, spatial locality, and orientation selectivity to cope with the variations due to illumination and facial expression changes. The kernel PCA method [42] is then extended to include fractional power polynomial models for enhanced face recognition performance. A fractional power polynomial, however, does not necessarily define a kernel function, as it might not define a positive semi-definite Gram matrix (see Sect. 3.4). Note that the sigmoid kernels (see Eq. 12), one of the three classes of widely used kernel functions (polynomial kernels, Gaussian kernels, and sigmoid kernels), do not actually define a positive semi-definite Gram matrix, either [42]. Nevertheless, the sigmoid kernels have been successfully used in practice, such as in building support vector machines. In order to derive real kernel PCA features, we apply only those kernel PCA eigenvectors that are associated with positive eigenvalues.

The feasibility of the Gabor-based kernel PCA method with fractional power polynomial models has been successfully tested on both frontal and pose-angled face recognition, using two data

sets from the FERET database [38] and the CMU PIE database [44], respectively. The FERET data set contains 600 frontal face images of 200 subjects, while the PIE data set consists of 680 images across 5 poses (left and right profiles, left and right half profiles, and frontal view) with 2 different facial expressions (neutral and smiling) of 68 subjects. The effectiveness of the Gabor-based kernel PCA method with fractional power polynomial models is shown in terms of both absolute performance indices and comparative performance against the PCA method, the kernel PCA method with polynomial kernels, the kernel PCA method with fractional power polynomial models, the Gabor wavelet based PCA method, and the Gabor wavelet based kernel PCA method with polynomial kernels.

## 2 Background

Robust face recognition schemes require both low dimensional feature representation for data compression purposes and enhanced discrimination abilities for subsequent image retrieval. The representation methods usually start with a dimensionality reduction procedure, since the high-dimensionality of the original visual space makes the statistical estimation very difficult, if not impossible, due to the fact that the high-dimensional space is mostly empty. Popular representation methods for face recognition include Principal Component Analysis (PCA) [22], [47], [26], shape and texture ('shape-free' image) of faces [9], [4], [49], [24], [27], and Gabor wavelet representation [12], [31], [23], [50], [28]. The discrimination methods often try to achieve the purpose of high separability between the different patterns in whose classification one is interested [26], [25]. Commonly used discrimination methods include Bayes classifier and the MAP rule [33], [28], Fisher Linear Discriminant (FLD) [46], [3], [17], [27], and more recently kernel PCA method [43], [42], [51], [32].

Gabor wavelets model quite well the receptive field profiles of cortical simple cells [12], [13]. The Gabor wavelet representation, therefore, captures salient visual properties such as spatial localization, orientation selectivity, spatial frequency characteristic. Lades et al. [23] applied Gabor wavelets for face recognition using the Dynamic Link Architecture (DLA) framework. The DLA starts by computing the Gabor jets, and then it performs a flexible template comparison between

the resulting image decompositions using graph-matching. Wiskott et al. [50] have expanded on DLA when they developed a Gabor wavelet based elastic bunch graph matching method to label and recognize human faces. Based on the 2D Gabor wavelet representation and the labeled elastic graph matching, Lyons et al. [30], [29] proposed an algorithm for two-class categorization of gender, race, and facial expression. Recently, Donato et al. [16] have compared a method based on Gabor representation with other techniques and found that the former gave better performance.

Kernel PCA overcomes many limitations of its linear counterpart by nonlinearly mapping the input space to a high-dimensional feature space. Being linear in the feature space, but nonlinear in the input space, kernel PCA thus is capable of deriving low-dimensional features that incorporate higher order statistics. The underlying justification of kernel PCA comes from Cover's theorem on the separability of patterns, which states that nonlinearly separable patterns in an input space are linearly separable with high probability if the input space is transformed nonlinearly to a high dimensional feature space [20]. Computationally, kernel PCA takes advantage of the Mercer equivalence condition and is feasible because the dot products in the high dimensional feature space are replaced by those in the input space while computation complexity is related to the number of training examples rather than the dimension of the feature space. Scholkopf et al. [43] showed that kernel PCA outperforms PCA using an adequate non-linear representation of input data. Yang et al. [51] compared face recognition performance using kernel PCA and the Eigenfaces method, and their results showed that the kernel PCA method with a cubic polynomial kernel achieved the lowest error rate. Moghaddam [32] demonstrated that kernel PCA with Gaussian kernels performed better than PCA for face recognition.

## 3 Gabor-based Kernel PCA with Fractional Power Polynomial Models for Face Recognition

This section details the novel Gabor-based kernel PCA method with fractional power polynomial models for face recognition. First, Gabor wavelet representation of face images derives desirable features characterized by spatial frequency, spatial locality, and orientation selectivity. Thus these features should be robust to variations due to illumination and facial expression changes [39], [34],

[41]. Second, kernel PCA works on the Gabor wavelet representation and nonlinearly derives low-dimensional features that incorporate higher order statistics. The computationally costly nonlinear mapping is never implemented explicitly by kernel PCA, which instead applies the kernel trick to achieve the same goal [42], [10]. Finally, the novel Gabor-based kernel PCA method applies fractional power polynomial models for face recognition.

### 3.1 Gabor Wavelets

Gabor wavelets were introduced to image analysis due to their biological relevance and computational properties [31], [13], [21], [14]. The Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains.

The Gabor wavelets (kernels, filters) can be defined as follows [12], [31], [23]:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}} \left[ e^{ik_{\mu,\nu} z} - e^{-\frac{\sigma^2}{2}} \right] \tag{1}$$

where $\mu$ and $\nu$ define the orientation and scale of the Gabor kernels, $z = (x, y)$, $\|\cdot\|$ denotes the norm operator, and the wave vector $k_{\mu,\nu}$ is defined as follows:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \tag{2}$$

where $k_\nu = k_{max}/f^\nu$ and $\phi_\mu = \pi\mu/8$. $k_{max}$ is the maximum frequency, and $f$ is the spacing factor between kernels in the frequency domain [23].

The Gabor kernels in Eq. 1 are all self-similar since they can be generated from one filter, the mother wavelet, by scaling and rotation via the wave vector $k_{\mu,\nu}$. Each kernel is a product of a Gaussian envelope and a complex plane wave, while the first term in the square brackets in Eq. 1 determines the oscillatory part of the kernel and the second term compensates for the DC value. The effect of the DC term becomes negligible when the parameter $\sigma$, which determines the ratio of the Gaussian window width to wavelength, has sufficiently large values. In most cases one would use Gabor wavelets of five different scales, $\nu \in \{0, ..., 4\}$, and eight orientations, $\mu \in \{0, ..., 7\}$ [18], [21], [7].

### 3.2 Gabor Feature Representation

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined by Eq. 1. Let $I(x, y)$ be the gray level distribution of an image, the convolution of image $I$ and a Gabor kernel $\psi_{\mu,\nu}$ is defined as follows:

$$O_{\mu,\nu}(z) = I(z) * \psi_{\mu,\nu}(z) \tag{3}$$

where $z = (x, y)$, $*$ denotes the convolution operator, and $O_{\mu,\nu}(z)$ is the convolution result corresponding to the Gabor kernel at orientation $\mu$ and scale $\nu$. Therefore, the set $\mathcal{S} = \{O_{\mu,\nu}(z) : \mu \in \{0, ..., 7\}, \nu \in \{0, ..., 4\}\}$ forms the Gabor wavelet representation of the image $I(z)$

To encompass different spatial frequencies (scales), spatial localities, and orientation selectivities, we concatenate all these representation results and derive an augmented feature vector $\mathcal{X}$. Before the concatenation, we first downsample each $O_{\mu,\nu}(z)$ by a factor $\rho$ to reduce the space dimension, and normalize it to zero mean and unit variance. We then construct a vector out of the $O_{\mu,\nu}(z)$ by concatenating its rows (or columns). Now, let $O_{\mu,\nu}^{(\rho)}$ denote the normalized vector constructed from $O_{\mu,\nu}(z)$ (downsampled by $\rho$ and normalized to zero mean and unit variance), the augmented Gabor feature vector $\mathcal{X}^{(\rho)}$ is then defined as follows:

$$\mathcal{X}^{(\rho)} = \left( O_{0,0}^{(\rho)^t} \quad O_{0,1}^{(\rho)^t} \quad \cdots \quad O_{4,7}^{(\rho)^t} \right)^t \tag{4}$$

where $t$ is the transpose operator. The augmented Gabor feature vector thus encompasses all the elements (downsampled and normalized) of the Gabor wavelet representation set, $\mathcal{S} = \{O_{\mu,\nu}(z) : \mu \in \{0, ..., 7\}, \nu \in \{0, ..., 4\}\}$, as important discriminating information.

### 3.3 Kernel PCA

Principal Component Analysis (PCA), a powerful technique for reducing a large set of correlated variables to a smaller number of uncorrelated components, has been applied extensively for both face representation and recognition. Kirby and Sirovich [22] showed that any particular face can be economically represented along the eigenpictures coordinate space, and that any face can be

approximately reconstructed by using just a small collection of eigenpictures and the corresponding projections. Applying PCA technique to face recognition, Turk and Pentland [47] developed a well-known Eigenfaces method, where the eigenfaces correspond to the eigenvectors associated with the largest eigenvalues of the face covariance matrix. The eigenfaces thus define a feature space, or "face space", which drastically reduces the dimensionality of the original space, and face detection and recognition are then carried out in the reduced space. Based on PCA, a host of face recognition methods have been developed to improve classification accuracy and generalization performance [23], [46], [3], [17], [50], [26], [28].

The PCA technique, however, encodes only for 2nd order statistics, namely, the variances and the covariances. As these 2nd order statistics provide only partial information on the statistics of both natural images and human faces, it might become necessary to incorporate higher order statistics as well. Towards that end, PCA is extended to a nonlinear form by mapping nonlinearly the input space to a feature space, where PCA is ultimately implemented. Due to the nonlinear mapping between the input space and the feature space, this form of PCA is nonlinear, and naturally called nonlinear PCA [20]. Applying different mappings, nonlinear PCA can encode arbitrary higher-order correlations among the input variables. The underlying justification of nonlinear PCA comes from Cover's theorem on the separability of patterns, which states that nonlinearly separable patterns in an input space are linearly separable with high probability if the input space is transformed nonlinearly to a high dimensional feature space [20]. While the Hebbian networks, the replicator networks, and the principal curves are all capable of implementing nonlinear PCA, kernel PCA enjoys simple implementation by means of kernel functions [15], [43]. The nonlinear mapping between the input space and the feature space, with a possibly prohibitive computational cost, is never implemented explicitly by kernel PCA [43]. Rather, kernel PCA applies kernel functions in the input space to achieve the same effect of the expensive nonlinear mapping. Specifically, kernel PCA takes advantage of the Mercer equivalence condition and is feasible because the dot products in the high dimensional feature space are replaced by those in the input space while computation is related to the number of training exemplars rather than the dimension of the feature space.

Let $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M \in \mathbb{R}^N$ be the data in the input space, and let $\Phi$ be a nonlinear mapping

between the input space and the feature space: $\Phi : \mathbb{R}^N \rightarrow F$. Note that for kernel PCA, the nonlinear mapping, $\Phi$, usually defines a kernel function [42]. Assume the mapped data is centered (see [43] for centering data in the feature space), and let $\mathcal{D}$ represent the data matrix in the feature space: $\mathcal{D} = [\Phi(\mathcal{X}_1)\Phi(\mathcal{X}_2)\cdots\Phi(\mathcal{X}_M)]$. Let $K \in \mathbb{R}^{M\times M}$ define a kernel matrix by means of dot product in the feature space:

$$K_{ij} = (\Phi(\mathcal{X}_i) \cdot \Phi(\mathcal{X}_j)) \tag{5}$$

Scholkopf [43] shows that the eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_M$, and the eigenvectors, $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M$, of kernel PCA can be derived by solving the following eigenvalue equation:

$$K\mathcal{A} = M\mathcal{A}\Lambda \quad with \;\; \mathcal{A} = [\alpha_1\alpha_2\ldots\alpha_M], \Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_M\} \tag{6}$$

where $\mathcal{A} \in \mathbb{R}^{M\times M}$ is an orthogonal eigenvector matrix, $\Lambda \in \mathbb{R}^{M\times M}$ a diagonal eigenvalue matrix with diagonal elements in decreasing order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$), and $M$ a constant — the number of training samples. In order to derive the eigenvector matrix, $\mathcal{V} = [\mathcal{V}_1\mathcal{V}_2\ldots\mathcal{V}_M]$, of kernel PCA, first, $\mathcal{A}$ should be normalized such that $\lambda_i\|\alpha_i\|^2 = 1, i = 1, 2, \ldots, M$. The eigenvector matrix, $\mathcal{V}$, is then derived as follows:

$$\mathcal{V} = \mathcal{D}\mathcal{A} \tag{7}$$

Let $\mathcal{X}$ be a test sample whose image in the feature space is $\Phi(\mathcal{X})$. The kernel PCA features of $\mathcal{X}$ are derived as follows:

$$\mathcal{F} = \mathcal{V}^t\Phi(\mathcal{X}) = \mathcal{A}^t\mathcal{B} \tag{8}$$

where $\mathcal{B} = [\Phi(\mathcal{X}_1) \cdot \Phi(\mathcal{X}) \;\; \Phi(\mathcal{X}_2) \cdot \Phi(\mathcal{X}) \;\cdots\; \Phi(\mathcal{X}_M) \cdot \Phi(\mathcal{X})]^t$.

### 3.4 Kernel PCA with Fractional Power Polynomial Models

The kernel matrix (see Eq. 5) and the kernel PCA features (see Eq. 8) are both defined on dot products in the high dimensional feature space, whose computation might be prohibitively expensive. Kernel PCA, however, manages to compute the dot products by means of a kernel function [43], [42]:

$$k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})) \tag{9}$$

This section first reviews the concept of Gram matrix and a sufficient and necessary condition for a symmetric function to be a kernel function [42], [10], and then presents kernel PCA with fractional power polynomial models.

**Definition 1** *Gram Matrix: Given a finite data set $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M\}$ in the input space and a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ (or $\mathbb{C}$), the $M \times M$ matrix K with elements $K_{ij} = k(\mathcal{X}_i, \mathcal{X}_j)$ is called Gram matrix of $k$ with respect to $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M$.*

**Definition 2** *Kernel Function: A sufficient and necessary condition for a symmetric function to be a kernel function is that its Gram matrix is positive semi-definite.*

Three classes of kernel functions widely used in kernel classifiers, neural networks, and support vector machines are polynomial kernels, Gaussian kernels, and sigmoid kernels [42]:

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d \tag{10}$$

$$k(\mathbf{x}, \mathbf{y}) = exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \tag{11}$$

$$k(\mathbf{x}, \mathbf{y}) = tanh(\kappa(\mathbf{x} \cdot \mathbf{y}) + \vartheta) \tag{12}$$

where $d \in \mathbb{N}$, $\sigma > 0$, $\kappa > 0$, and $\vartheta < 0$.

Note that the sigmoid kernels (see Eq. 12) do not actually define a positive semi-definite Gram matrix, hence are not kernel functions by definition (see Definition 2). Nevertheless, the sigmoid kernels have been successfully used in practice, such as in building support vector machines [42]. In this paper, we extend the polynomial kernels defined by Eq. 10 to include fractional power polynomial models, namely, $0 < d < 1$. In order to derive real kernel PCA features (Eq. 8), we apply only those kernel PCA eigenvectors that are associated with positive eigenvalues.

### 3.5 Similarity Measures and Classification Rule

The Gabor-based kernel PCA method integrates the Gabor wavelet representation of face images and kernel PCA with fractional power polynomial models for face recognition. When a face image is presented to the Gabor-based kernel PCA classifier, the augmented Gabor feature vector of the

image is first calculated as detailed in Sect. 3.2, and the low-dimensional Gabor-based kernel PCA features, $\mathcal{F}$, are derived using Eq. 8.

Let $\mathcal{M}_k^0$, $k = 1, 2, \ldots, L$, be the mean of the training samples for class $\omega_k$. The classifier applies, then, the nearest neighbor (to the mean) rule for classification using some similarity (distance) measure $\delta$:

$$\delta(\mathcal{F}, \mathcal{M}_k^0) = \min_j \delta(\mathcal{F}, \mathcal{M}_j^0) \quad \longrightarrow \quad \mathcal{F} \in \omega_k \tag{13}$$

The Gabor-based kernel PCA feature vector, $\mathcal{F}$, is classified as belonging to the class of the closest mean, $\mathcal{M}_k^0$, using the similarity measure $\delta$.

Popular similarity measures include $L_1$ distance measure, $\delta_{L_1}$, $L_2$ distance measure, $\delta_{L_2}$, Mahalanobis distance measure, $\delta_{Md}$, and cosine similarity measure, $\delta_{cos}$, which are defined as follows, respectively:

$$\delta_{L_1}(\mathcal{X}, \mathcal{Y}) = \sum_i |\mathcal{X}_i - \mathcal{Y}_i| \tag{14}$$

$$\delta_{L_2}(\mathcal{X}, \mathcal{Y}) = (\mathcal{X} - \mathcal{Y})^t (\mathcal{X} - \mathcal{Y}) \tag{15}$$

$$\delta_{Md}(\mathcal{X}, \mathcal{Y}) = (\mathcal{X} - \mathcal{Y})^t \Sigma^{-1} (\mathcal{X} - \mathcal{Y}) \tag{16}$$

$$\delta_{cos}(\mathcal{X}, \mathcal{Y}) = \frac{-\mathcal{X}^t \mathcal{Y}}{\|\mathcal{X}\| \, \|\mathcal{Y}\|} \tag{17}$$

where $\Sigma$ is the covariance matrix, and $\|\cdot\|$ denotes the norm operator. Note that the cosine similarity measure includes a minus sign in Eq. 17, because the nearest neighbor (to the mean) rule of Eq. 13 applies minimum (distance) measure rather than maximum similarity measure.

## 4 Experiments of Frontal and Pose-angled Face Recognition

This section assesses the performance of the Gabor-based kernel PCA method with fractional power polynomial models for both frontal and pose-angled face recognition. For frontal face recognition, the data set is from the FERET database [38], and it contains 600 frontal face images corresponding to 200 subjects. The images are acquired under variable illumination and facial expression. For pose-angled face recognition, the data set is from the PIE database [44], and it contains 680 images across 5 poses with 2 different facial expressions (neutral and smiling) of 68

subjects. The effectiveness of the Gabor-based kernel PCA method with fractional power polynomial models is shown in terms of both absolute performance indices and comparative performance against the PCA method, the kernel PCA method with polynomial kernels, the kernel PCA method with fractional power polynomial models, the Gabor wavelet based PCA method, and the Gabor wavelet based kernel PCA method with polynomial kernels.

### 4.1 Frontal Face Recognition Performance of the Gabor-based Kernel PCA Method with Fractional Power Polynomial Models

The FacE REcognition Technology (FERET) facial database [38] displays diversity across gender, ethnicity, and age. Since images are acquired during different photo sessions, the illumination condition, facial expression, and the size of the face may vary. The FERET database has become the de facto standard for evaluating face recognition technologies. The data set used in our experiments consists of 600 FERET frontal face images corresponding to 200 subjects, such that each subject has 3 images of size $256 \times 384$ with 256 gray scale levels. Data preparation then normalizes the face images and extracts facial regions that contain only face, so that the performance of face recognition is not affected by the factors not related to face, such as hair styles. Specifically, the normalization consists of the following procedures: first, manual annotation detects the centers of the eyes; second, rotation and scaling transformations align the centers of the eyes to predefined locations and fixed interocular distance; finally, a subimage procedure crops the face image to the size of $128 \times 128$ to extract the facial region.

Fig. 1 shows some example FERET images used in our experiments that are already cropped to the size of $128 \times 128$ to extract the facial region. Note that each subject has three images, which are acquired during different photo sessions under variable illumination and facial expression. As two images for each subject are randomly chosen for training and the remaining image (unseen during training) is used for testing (see Fig. 1), the training set includes 400 images while the testing set has 200 images. The Gabor-based kernel PCA method thus has to cope with both illumination and facial expression variability.

For comparison purpose, the first set of experiments implements the kernel PCA method using

a polynomial kernel with degree one, i.e. $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$. This special case of kernel PCA is equivalent to PCA [42]. Fig. 2 shows face recognition performance of the kernel PCA method using the four different similarity measures: the $L_1$ distance measure, the $L_2$ distance measure, the Mahalanobis distance measure, and the cosine similarity measure. The horizontal axis indicates the number of features used, and the vertical axis represents the correct face recognition rate, which is the rate that the top response is correct (in the correct class). The Mahalanobis distance measure performs the best, followed in order by the $L_1$ distance measure, the $L_2$ distance measure, and the cosine similarity measure. The reason for such an ordering is that the Mahalanobis distance measure counteracts the fact that $L_1$ and $L_2$ distance measures in the PCA space weight preferentially for low frequencies. These results are consistent with those reported by Moghaddam and Pentland [33] and Sung and Poggio [45]. As the $L_2$ distance measure weights more the low frequencies than $L_1$ does, the $L_1$ distance measure should perform better than the $L_2$ distance measure, a conjecture validated by our experiments. The cosine similarity measure does not compensate the low frequency preference, and it performs the worst among all the measures. The experimental results provide a baseline face recognition performance based on the intensity images, and suggest that one should use the Mahalanobis distance measure for the following comparative assessment of different face recognition methods.

The second set of experiments assesses face recognition performance of the kernel PCA method with polynomial kernels, $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d, d \in \mathbb{N}$. Fig. 3 shows face recognition performance of the kernel PCA method with four different degrees of polynomial kernels using the Mahalanobis distance measure: $d = 1$, $d = 2$, $d = 3$, and $d = 4$. The first order polynomial ($d = 1$) kernel PCA method performs the best, followed in order by the second order polynomial ($d = 2$) kernel PCA method, the third order polynomial ($d = 3$) kernel PCA method, and the forth order polynomial ($d = 4$) kernel PCA method. Note that the first order polynomial kernel PCA method is equivalent to the PCA method [42], hence the results show that the PCA method performs better than the kernel PCA method with the second, the third, or the fourth order polynomial kernel. Fig. 3 also shows that among the four different degrees of polynomial kernels, the lower the degree is the better the kernel PCA method performs. It thus seems natural that the kernel PCA method

should be extended to include some fractional power polynomials, whose degrees are even lower than one, i.e. $0 < d < 1$, in order to achieve better face recognition performance. A fractional power polynomial, however, is not necessarily a kernel, as it might not define a positive semi-definite Gram matrix as detailed in Sect. 3.4. Thus the fractional power polynomials are called models rather than kernels in this paper. Note that the sigmoid kernels (see Eq. 12), one of the three classes of widely used kernel functions, do not actually define a positive semi-definite Gram matrix, either [42]. Nevertheless, the sigmoid kernels have been successfully used in practice, such as in building support vector machines. In this paper, we extend the polynomial kernels defined by Eq. 10 to include fractional power polynomial models, namely, $0 < d < 1$. In order to derive real kernel PCA features, we apply only those kernel PCA eigenvectors that are associated with positive eigenvalues.

The third set of experiments evaluates the kernel PCA method with fractional power polynomial models, $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d, 0 < d < 1$. Fig. 4 shows face recognition performance of the kernel PCA method with three fractional power polynomial models using the Mahalanobis distance measure: Md_0.6 ($d = 0.6$), Md_0.7 ($d = 0.7$), and Md_0.8 ($d = 0.8$). The top three curves correspond to the face recognition performance of the kernel PCA method with these three different fractional power polynomial models. The bottom curve is the face recognition performance of the kernel PCA method with the kernel function $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$. Fig. 4 and Fig. 3 show that the kernel PCA method with fractional power polynomial models performs better than the kernel PCA method with polynomial kernels.

To further improve face recognition performance, we combine the Gabor wavelet representation and the kernel PCA method. In particular, the fourth set of experiments implements the Gabor wavelet based kernel PCA method with polynomial kernels, $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d, d \in \mathbb{N}$. Fig. 5 shows face recognition performance of the Gabor wavelet based kernel PCA method with four different degrees of polynomial kernels using the Mahalanobis distance measure: $d = 1$, $d = 2$, $d = 3$, and $d = 4$. In comparison, Fig. 5 and Fig. 3 show that the Gabor wavelet based kernel PCA method performs better than the kernel PCA method by large margins. Fig. 5 also shows that the face recognition performance of the Gabor wavelet based kernel PCA method decreases as the

degree of the polynomial kernel increases, a result consistent with the one derived by the kernel PCA method with polynomial kernels (see Fig. 3).

The next set of experiments assesses face recognition performance of the Gabor-based kernel PCA method with fractional power polynomial models, $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d, 0 < d < 1$. Fig. 6 shows face recognition performance of the Gabor wavelet based kernel PCA method with a fractional power polynomial model using the Mahalanobis distance measure: Md_Gabor_0.6 ($d = 0.6$). Note that the performance curves of the kernel PCA method (Md_1) and the Gabor wavelet based kernel PCA method (Md_Gabor_1), with the kernel function $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$, are also included for comparison. The experimental results show that the Gabor-based kernel PCA method with fractional power polynomial models performs better than the Gabor wavelet based kernel PCA method with polynomial kernels, followed by the kernel PCA method with polynomial kernels. In particular, the Gabor-based kernel PCA method with a fractional power polynomial model ($d = 0.6$) achieves 99.5% correct face recognition accuracy using 246 features.

### 4.2 Pose-angled Face Recognition Performance of the Gabor-based Kernel PCA Method with Fractional Power Polynomial Models

The CMU Pose, Illumination, and Expression (PIE) database [44] contains over 40,000 facial images of 68 people. The images are acquired across different poses, under variable illumination conditions, and with different facial expressions [44]. The data set used in our experiments consists of 680 images across 5 pose classes (left and right profiles, left and right half profiles, and frontal view) with 2 different facial expressions (neutral and smiling) of the 68 subjects, such that each subject has 10 images of size 640 $\times$ 486 with 256 gray scale levels. For each subject, one of the two facial expressions is randomly chosen for training, with the remaining one for testing. The training set thus contains 340 images across 5 different poses, such that each pose class has 68 images corresponding to the 68 different subjects. The testing set contains the remaining 340 images with each pose having 68 images of the 68 subjects, respectively. Data preparation then normalizes the face images and extracts facial regions that contain only face. Fig. 7 shows some example PIE face images used in our experiments that are already cropped to the size of 128 $\times$ 128 to extract the

facial region. In particular, the top row shows images used in training and the bottom row shows images for testing. In each row, there are 5 images corresponding to the 5 different poses: the right profile, the right half profile, the frontal view, the left half profile, and the left profile.

There are two major approaches to address pose-angled face recognition: 3D model-based methods and 2D view-based methods. 3D model-based face recognition methods account for variations in pose by applying 3D models either derived from laser-scanned 3D heads (range data) or reconstructed from 2D images [1], [19], [5]. 2D view-based face recognition methods identify the pose parameter by classifying face images into different pose classes, such as the parametric manifold formed by different views [2] and the multiple view-based and modular eigenspaces [37].

Our 2D view-based pose-angled face recognition method consists of two major procedures, namely the pose classification procedure and the face recognition procedure, which are carried out sequentially. The pose classification procedure, which applies a nearest neighbor (to the mean) classifier using the cosine similarity measure, works in a low-dimensional PCA space. The face recognition procedure, which applies the Gabor-based kernel PCA method with fractional power polynomial models, operates in each individual pose class. Fig. 8 shows the system architecture of our pose-angled face recognition method. The top face image is a test image and NNC denotes a nearest neighbor to the mean classifier. The five images in the middle are the mean faces of the five pose classes derived from the training data, while the Gabor-based kernel PCA method with fractional power polynomial (FPP) models is the classifier for face recognition within each individual face class. The bottom face images are the training images in the pose class (the left half profile pose class). Note that the bold lines indicate the face recognition process across pose: when a test face image is presented to the classifier, the pose classification procedure first assigns it to the left half profile pose class using a nearest neighbor to the mean classifier (NNC), the face recognition procedure then matches this unknown test image to a training face within the pose class using the Gabor-based kernel PCA method with fractional power polynomial models.

Pose classification, which applies a nearest neighbor (to the mean) classifier, is assessed using 3 different similarity measures: the $L_1$ distance measure, the $L_2$ distance measure, and the cosine similarity measure, as defined in Sect. 3.5. Fig. 9 shows pose classification results using the cosine,

the $L_1$, and the $L_2$ similarity measure, where the means of the five pose classes are derived from the training data. The horizontal axis indicates the number of features used by the classifier, and the vertical axis the correct pose classification rate, which is the rate that the top response is the correct pose. The top curve shows face pose classification performance using the cosine similarity measure, followed by the two curves corresponding to the $L_1$ and the $L_2$ distance measure, respectively. In particular, face pose classification reaches 99.7% accuracy using only 23 features for the cosine similarity measure (the curves are drawn using an incremental step of 2, which show that the 99.7% accuracy occurs at 24 features). These experimental results show that the cosine similarity measure is most effective for pose classification. The reason for this finding is largely due to the different scales in the five pose classes. The $L_1$ and the $L_2$ distance measures are affected considerably by these different scales, as they involve the magnitude of two vectors (see Eqs. 14 and 15). The cosine similarity measure, however, calculates the angle between two vectors, and is not affected by their magnitudes (see Eq. 17), hence it performs better than either $L_1$ or $L_2$ distance measure for face pose classification.

Fig. 10 shows the pose-angled face recognition performance of the Gabor-based kernel PCA method with four polynomial kernels of degree 1 (Md_Gabor_1), degree 2 (Md_Gabor_2), degree 3 (Md_Gabor_3), and degree 4 (Md_Gabor_4), respectively, and the Gabor-based kernel PCA method with two fractional power polynomial models of degree 0.6 (Md_Gabor_0.6) and degree 0.7 (Md_Gabor_0.7), respectively, using the Mahalanobis distance measure. The experimental results evaluate and validate the feasibility of the Gabor-based kernel PCA method with fractional power polynomial models for pose-angled face recognition. Specifically, the Gabor-based kernel PCA method with fractional power polynomial models achieves the best face recognition performance across pose, followed in order by the kernel PCA method with polynomial kernels of degree 1, degree 2, degree 3, and degree 4, respectively. In particular, the Gabor-based kernel PCA method with a fractional power polynomial model of degree 0.7 achieves 95.3% pose-angled face recognition accuracy using 64 features.

## 5 Conclusions

This paper introduces a novel Gabor-based kernel PCA method with fractional power polynomial models for frontal and pose-angled face recognition. Gabor wavelets first derive desirable facial features characterized by spatial frequency, spatial locality, and orientation selectivity to cope with the variations due to illumination and facial expression changes. The kernel PCA method is then extended to include fractional power polynomial models for enhanced face recognition performance. The feasibility of the Gabor-based kernel PCA method with fractional power polynomial models has been successfully tested on both frontal and pose-angled face recognition, using two data sets from the FERET database and the CMU PIE database, respectively. The FERET data set contains 600 frontal face images of 200 subjects, while the PIE data set consists of 680 images across 5 poses (left and right profiles, left and right half profiles, and frontal view) with 2 different facial expressions (neutral and smiling) of 68 subjects. Experimental results show the effectiveness of the Gabor-based kernel PCA method with fractional power polynomial models for both frontal and pose-angled face recognition.

**Acknowledgments:** The author would like to thank the anonymous reviewers for their critical and constructive comments and suggestions. This work was partially supported by the TSWG R&D Contract N41756-03-C-4026.

## References

[1] J. J. Atick, P. A. Griffin, and A. N. Redlich, "Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images," *Neural Computation*, vol. 8, no. 6, pp. 1321–1340, 1996.

[2] S. Baker, S. K. Nayar, and H. Murase, "Parametric feature detection," *International Journal of Computer Vision*, vol. 27, no. 1, pp. 27–50, 1998.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[4] D. Beymer, "Vectorizing face images by interleaving shape and texture computations," A.I. memo No. 1537, Artificial Intelligence Laboratory, MIT, 1995.

[5] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[6] R. Brunelli and T. Poggio, "Face recognition: Features vs. templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1053, 1993.

[7] D. Burr, M. Morrone, and D. Spinelli, "Evidence for edge and bar detectors in human vision," *Vision Research*, vol. 29, no. 4, pp. 419–431, 1989.

[8] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[9] I. Craw and D. Tock, "The computer understanding of faces," in *Processing Images of Faces*, V. Bruce and M. Burton, Eds. Ablex Publishing Corporation, 1992.

[10] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other kernel-based learning methods*, Cambridge University Press, 2000.

[11] J. Daugman, "Face and gesture recognition: Overview," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 675–676, 1997.

[12] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research*, vol. 20, pp. 847–856, 1980.

[13] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional cortical filters," *Journal Opt. Soc. Amer.*, vol. 2, no. 7, pp. 1160–1169, 1985.

[14] J. G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1169–1179, 1988.

[15] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*, John Wiley, NY, 1996.

[16] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–989, 1999.

[17] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Am. A*, vol. 14, pp. 1724–1733, 1997.

[18] D. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[19] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[20] S. Haykin, *Neural Networks — A Comprehensive Foundation*, Prentice Hall, second edition, 1999.

[21] J. Jones and L. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J. Neurophysiology*, pp. 1233–1258, 1987.

[22] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, 1990.

[23] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Computers*, vol. 42, pp. 300–311, 1993.

[24] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743–756, 1997.

[25] C. Liu, "A Bayesian discriminating features method for face detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 725–740, 2003.

[26] C. Liu and H. Wechsler, "Evolutionary pursuit and its application to face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570–582, 2000.

[27] C. Liu and H. Wechsler, "A shape and texture based enhanced fisher classifier for face recognition," *IEEE Trans. on Image Processing*, vol. 10, no. 4, pp. 598–608, 2001.

[28] C. Liu and H. Wechsler, "Independent component analysis of Gabor features for face recognition," *IEEE Trans. on Neural Networks*, vol. 14, no. 4, pp. 919–928, 2003.

[29] M. J. Lyons, J. Budynek, , A. Plante, and S. Akamatsu, "Classifying facial attributes using a 2-d gabor wavelet representation and discriminant analysis," in *Proc. the fourth IEEE internatinoal conference on automatic face and gestrure recognition*, 2000.

[30] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999.

[31] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *Journal Opt. Soc. Amer.*, vol. 70, pp. 1297–1300, 1980.

[32] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 780–788, 2002.

[33] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, 1997.

[34] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 13, pp. 607–609, 1996.

[35] S. Pankanti, R. M. Bolle, and A. Jain, "Guest editors' introduction: Biometrics-the future of identification," *Computer*, vol. 33, no. 2, pp. 46–49, 2000.

[36] A. Pentland and T. Choudhury, "Face recognition for smart environments," *Computer*, vol. 33, no. 2, pp. 50–55, 2000.

[37] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. Computer Vision and Pattern Recognition*, 1994, pp. 84–91.

[38] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, pp. 295–306, 1998.

[39] R. Rao and D. Ballard, "An active vision architecture based on iconic representations," *Artificial Intelligence*, vol. 78, pp. 461–505, 1995.

[40] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expression: A survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65–77, 1992.

[41] B. Schiele and J. L. Crowley, "Recognition without correspondence using multidimensional receptive field histograms," *International Journal of Computer Vision*, vol. 36, no. 1, pp. 31–52, 2000.

[42] B. Scholkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press, 2002.

[43] B. Scholkopf, A. Smola, and K. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299–1319, 1998.

[44] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) database," in *Proc. Fifth International Conference on Automatic Face and Gesture Recognition*, Washington D. C., May, 2002.

[45] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39–51, 1998.

[46] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831–836, 1996.

[47] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 13, no. 1, pp. 71–86, 1991.

[48] D. Valentin, H. Abdi, A. J. O'Toole, and G. W. Cottrell, "Connectionist models of face processing: A survey," *Pattern Recognition*, vol. 27, no. 9, pp. 1209–1230, 1994.

[49] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 733–742, 1997.

[50] L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[51] M. H. Yang, N. Ahuja, and D. Kriegman, "Face recognition using kernel eigenfaces," in *Proc. IEEE International Conference on Image Processing*, Vancouver, Canada, September, 2000.

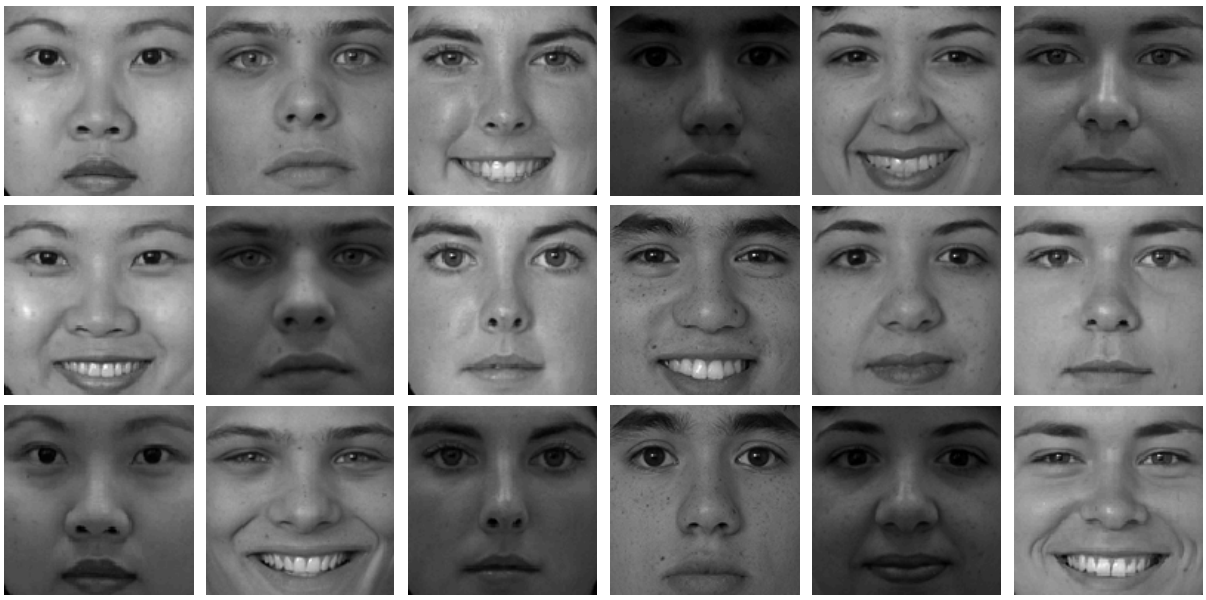

**Figure 1. Example FERET face images used in our experiments (cropped to the size of 128 $\times$ 128 to extract the facial region). Note that the images are acquired during different photo sessions under variable illumination and facial expression. Each subject has three images, two of which are randomly chosen for training while the remaining one (unseen during training) is for testing. Specifically, the top two rows show the examples of the training images while the bottom row shows the examples of the testing images.**

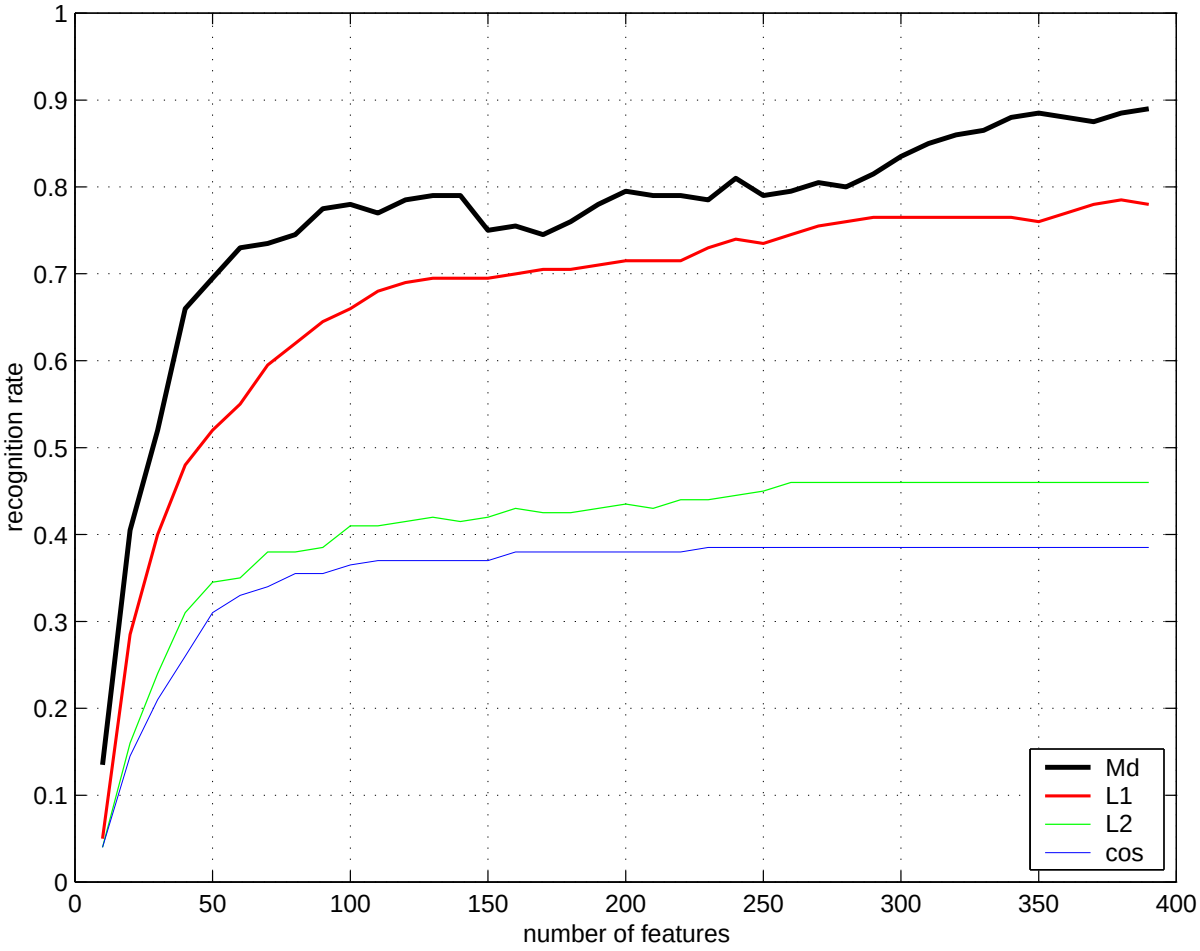


**Figure 2. Face recognition performance of the kernel PCA method with the kernel function $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$ using the four different similarity measures: Md (the Mahalanobis distance measure), L1 (the $L_1$ distance measure), L2 (the $L_2$ distance measure), and cos (the cosine similarity measure).**

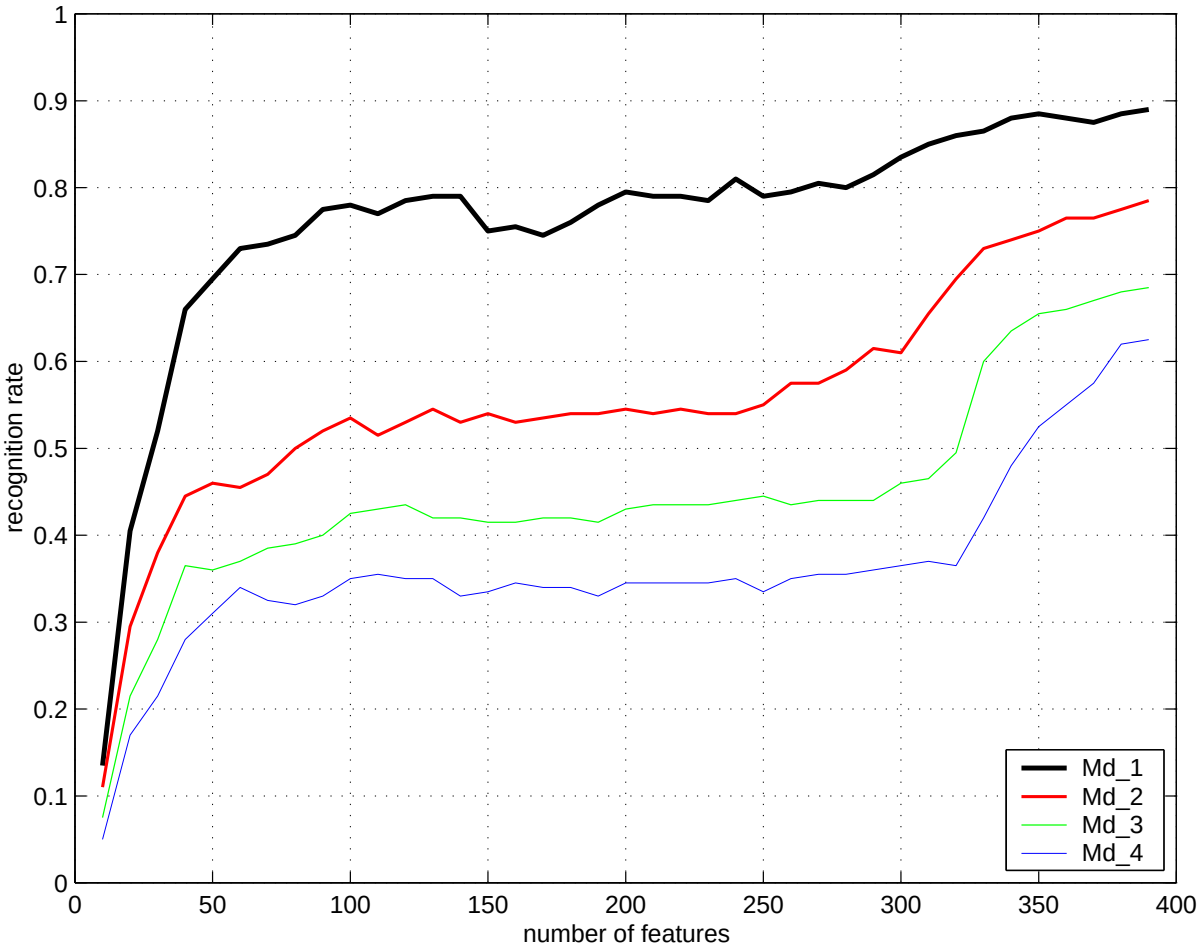


**Figure 3. Face recognition performance of the kernel PCA method with four different degrees of polynomial kernels using the Mahalanobis distance measure: Md_1 ($d = 1$), Md_2 ($d = 2$), Md_3 ($d = 3$), and Md_4 ($d = 4$).**

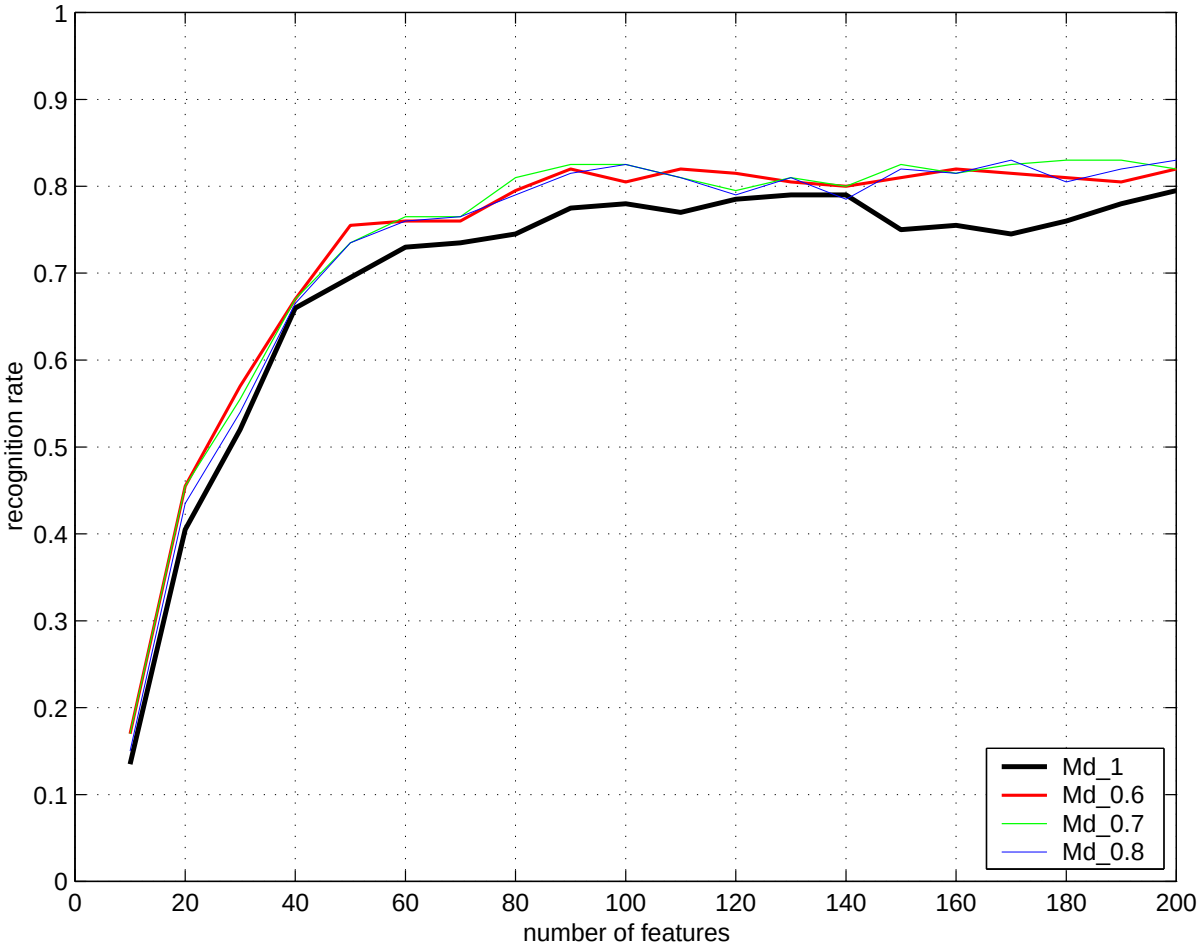


**Figure 4. Face recognition performance of the kernel PCA method with three fractional power polynomial models using the Mahalanobis distance measure: Md_0.6 ($d = 0.6$), Md_0.7 ($d = 0.7$), and Md_0.8 ($d = 0.8$). Note that the performance curve of the kernel PCA method (Md_1) with the kernel function $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$ is also included for comparison.**

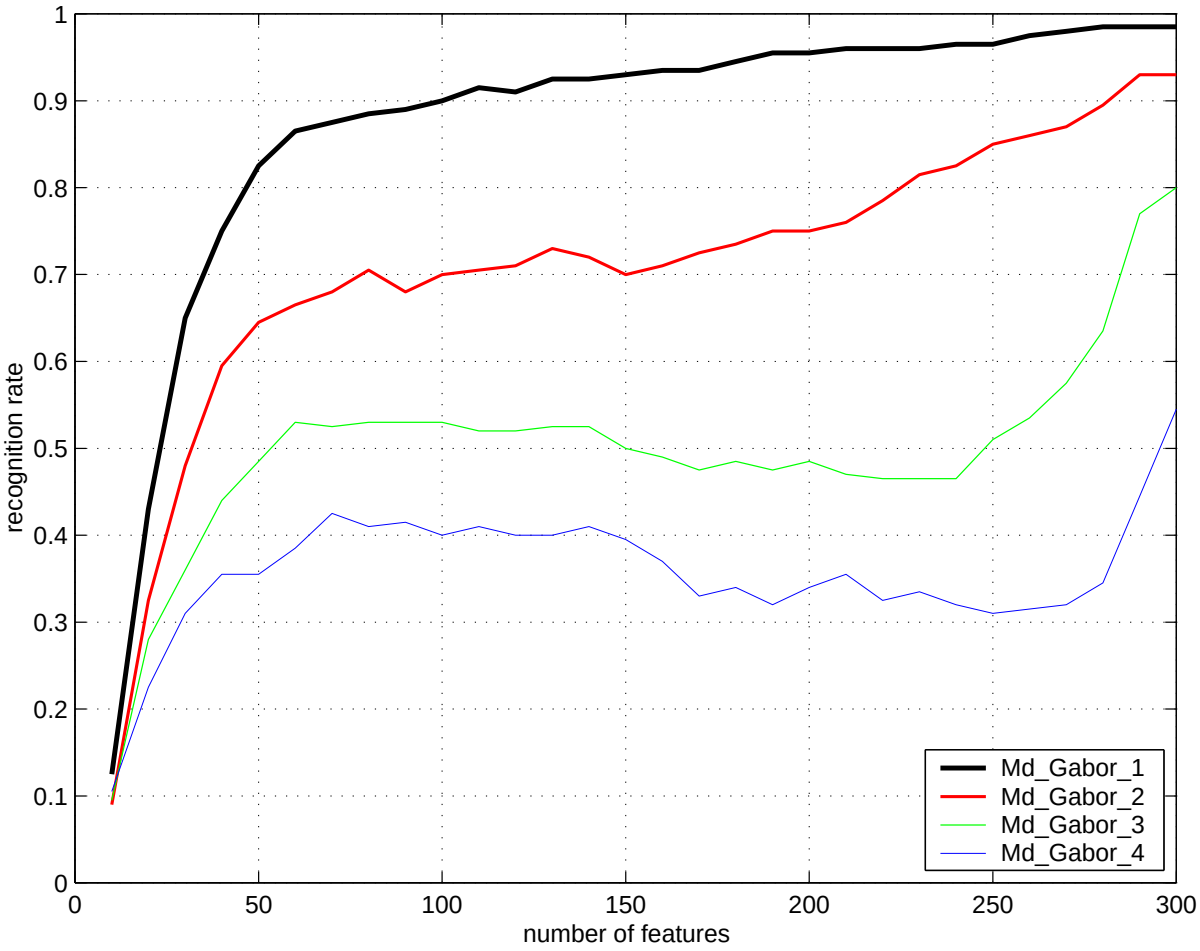


**Figure 5. Face recognition performance of the Gabor wavelet based kernel PCA method with four different degrees of polynomial kernels using the Mahalanobis distance measure: Md_Gabor_1 ($d = 1$), Md_Gabor_2 ($d = 2$), Md_Gabor_3 ($d = 3$), and Md_Gabor_4 ($d = 4$).**

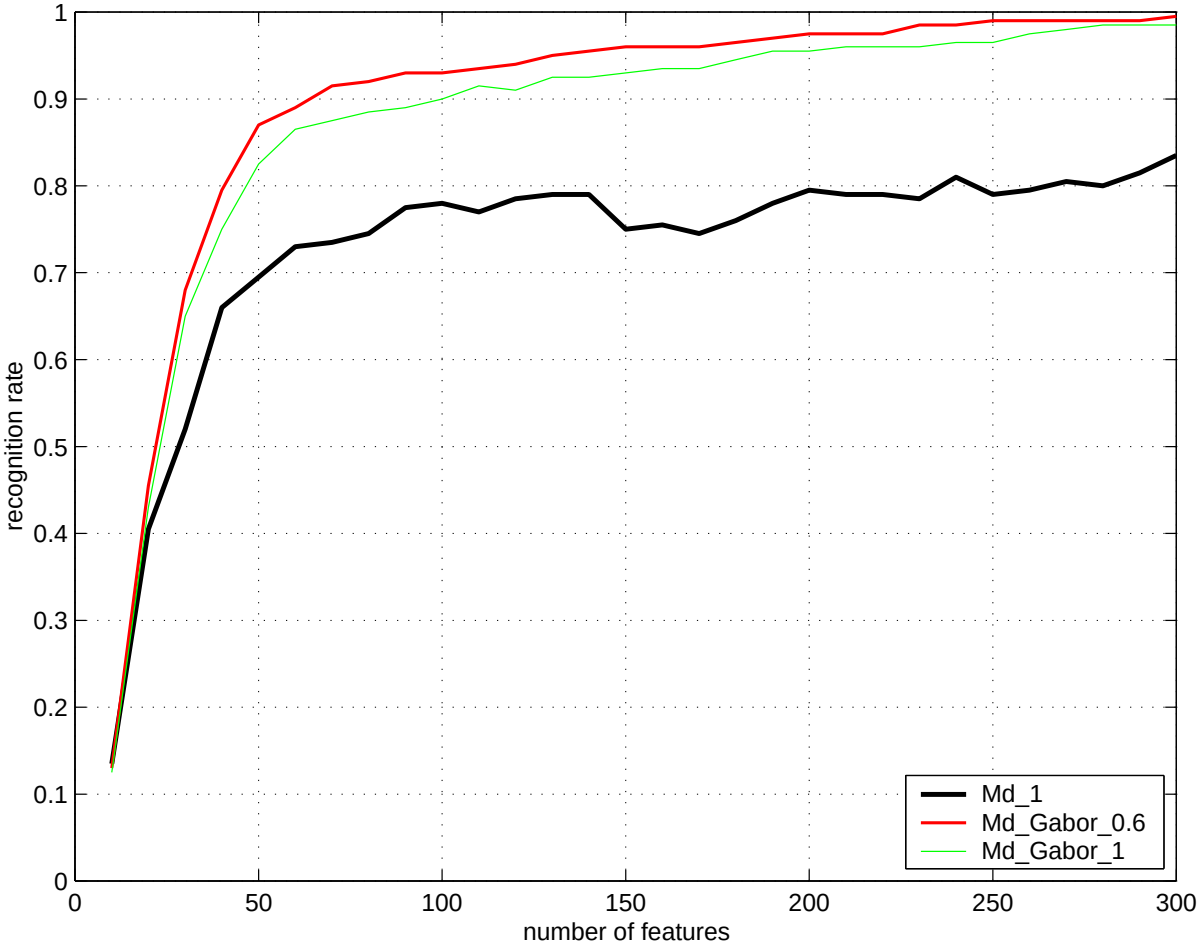


**Figure 6. Face recognition performance of the Gabor wavelet based kernel PCA method with a fractional power polynomial model using the Mahalanobis distance measure: Md_Gabor_0.6 ($d = 0.6$). Note that the performance curves of the kernel PCA method (Md_1) and the Gabor wavelet based kernel PCA method (Md_Gabor_1), with the kernel function $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$, are also included for comparison.**

**Figure 7. Example PIE face images used in our experiments (cropped to the size of 128 $\times$ 128 to extract the facial region). The top row shows images used in training and the bottom row shows images for testing. In each row, there are 5 images, which correspond to the 5 different poses: the right profile, the right half profile, the frontal view, the left half profile, and the left profile.**

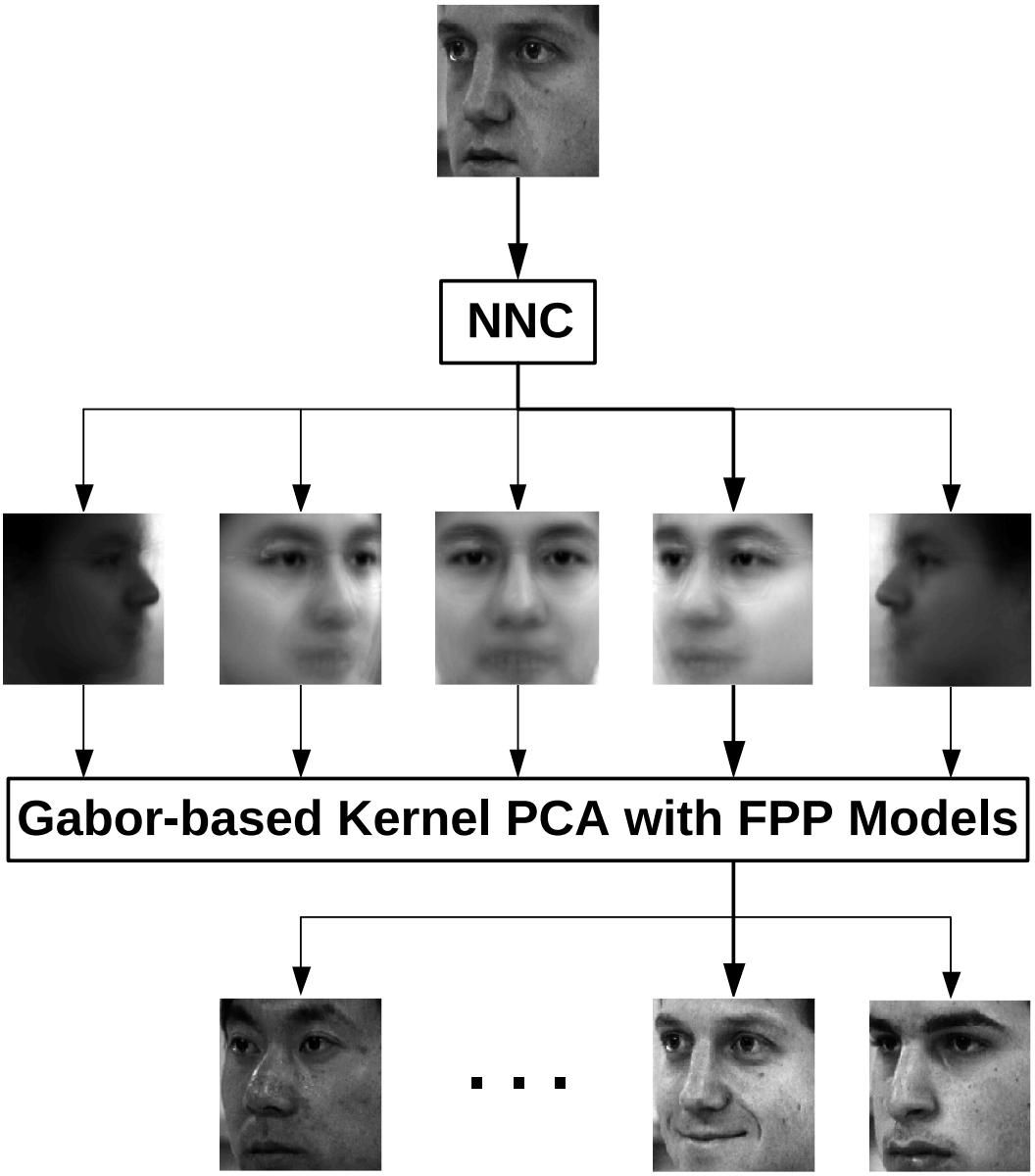


**Figure 8. The system architecture of the pose-angled face recognition method. The top face image is a test image and NNC denotes a nearest neighbor to the mean classifier. The five images in the middle are the mean faces of the five pose classes derived from the training data, while the Gabor-based kernel PCA method with fractional power polynomial (FPP) models is the classifier for face recognition within each individual face class. The bottom face images are the training images in the pose class (the left half profile pose class). Note that the bold lines indicate the face recognition process across pose.**

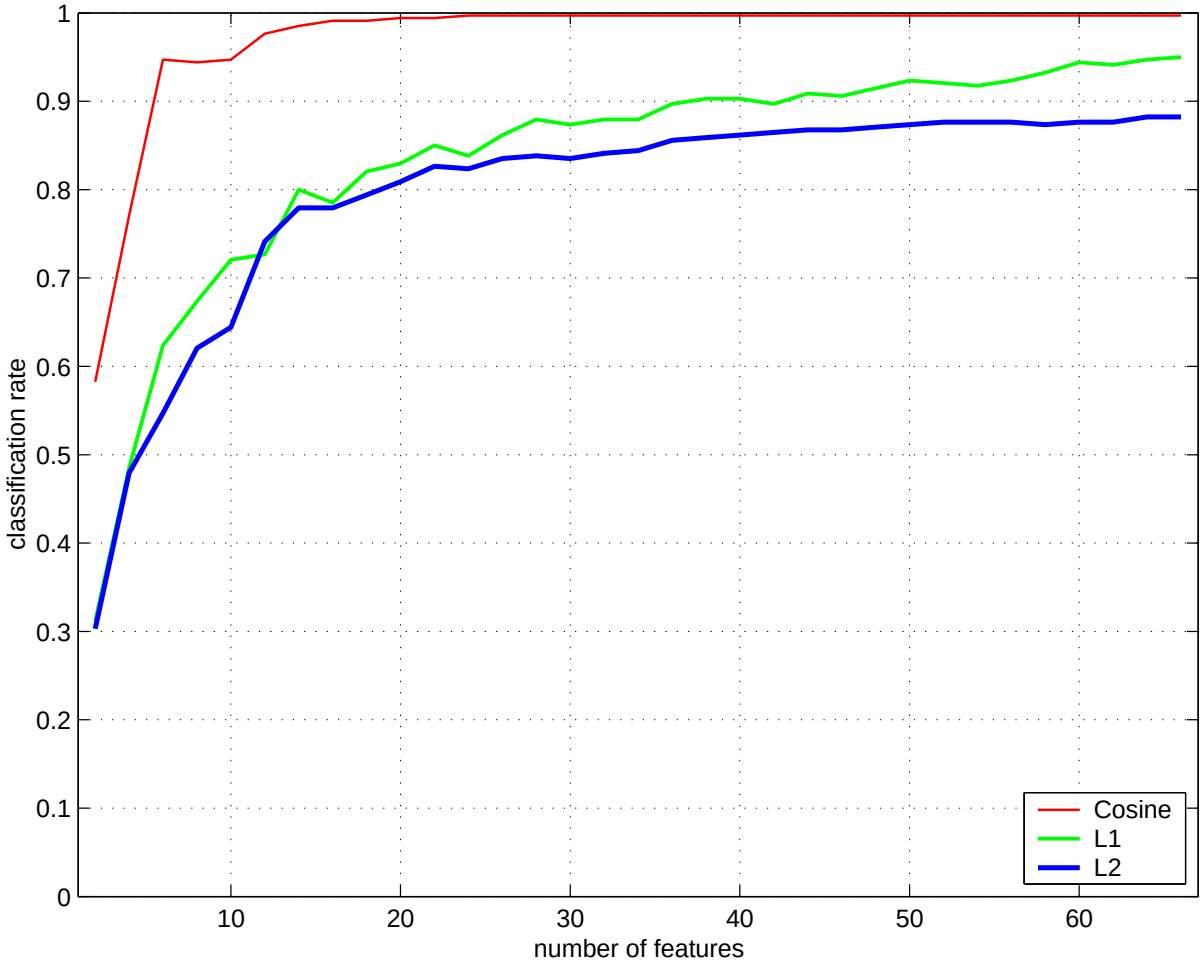


**Figure 9. Pose classification results using the cosine (Cosine), the $L_1$ (L1), and the $L_2$ (L2) similarity measure, where the means of the five pose classes are derived from the training data.**

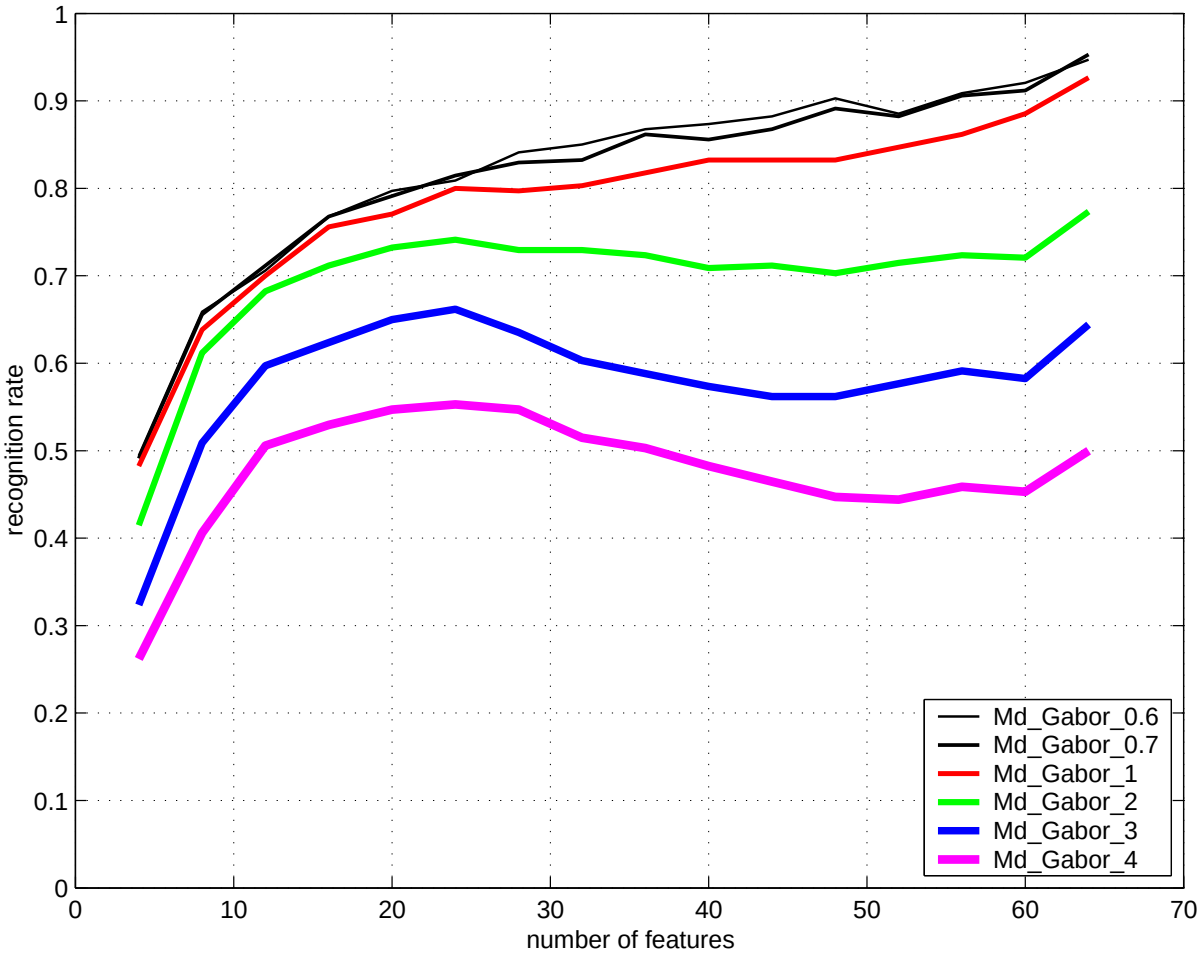


**Figure 10. Pose-angled face recognition performance of the Gabor-based kernel PCA method with four polynomial kernels of degree 1 (Md_Gabor_1), degree 2 (Md_Gabor_2), degree 3 (Md_Gabor_3), and degree 4 (Md_Gabor_4), respectively, and the Gabor-based kernel PCA method with two fractional power polynomial models of degree 0.6 (Md_Gabor_0.6) and degree 0.7 (Md_Gabor_0.7), respectively, using the Mahalanobis distance measure.**

## List of Figures


1 Example FERET face images used in our experiments (cropped to the size of 128 $\times$ 128 to extract the facial region). Note that the images are acquired during different photo sessions under variable illumination and facial expression. Each subject has three images, two of which are randomly chosen for training while the remaining one (unseen during training) is for testing. Specifically, the top two rows show the examples of the training images while the bottom row shows the examples of the testing images. . . . . 24

2 Face recognition performance of the kernel PCA method with the kernel function $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$ using the four different similarity measures: Md (the Mahalanobis distance measure), L1 (the $L_1$ distance measure), L2 (the $L_2$ distance measure), and cos (the cosine similarity measure). . . . . 25

3 Face recognition performance of the kernel PCA method with four different degrees of polynomial kernels using the Mahalanobis distance measure: Md_1 ($d = 1$), Md_2 ($d = 2$), Md_3 ($d = 3$), and Md_4 ($d = 4$). . . . . 26

4 Face recognition performance of the kernel PCA method with three fractional power polynomial models using the Mahalanobis distance measure: Md_0.6 ($d = 0.6$), Md_0.7 ($d = 0.7$), and Md_0.8 ($d = 0.8$). Note that the performance curve of the kernel PCA method (Md_1) with the kernel function $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$ is also included for comparison. . . . . 27

5 Face recognition performance of the Gabor wavelet based kernel PCA method with four different degrees of polynomial kernels using the Mahalanobis distance measure: Md_Gabor_1 ($d = 1$), Md_Gabor_2 ($d = 2$), Md_Gabor_3 ($d = 3$), and Md_Gabor_4 ($d = 4$). . . . . 28

6 Face recognition performance of the Gabor wavelet based kernel PCA method with a fractional power polynomial model using the Mahalanobis distance measure: Md_Gabor_0.6 ($d = 0.6$). Note that the performance curves of the kernel PCA method (Md_1) and the Gabor wavelet based kernel PCA method (Md_Gabor_1), with the kernel function $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$, are also included for comparison. . . . . 29

7 Example PIE face images used in our experiments (cropped to the size of 128 $\times$ 128 to extract the facial region). The top row shows images used in training and the bottom row shows images for testing. In each row, there are 5 images, which correspond to the 5 different poses: the right profile, the right half profile, the frontal view, the left half profile, and the left profile. . . . . 30

8 The system architecture of the pose-angled face recognition method. The top face image is a test image and NNC denotes a nearest neighbor to the mean classifier. The five images in the middle are the mean faces of the five pose classes derived from the training data, while the Gabor-based kernel PCA method with fractional power polynomial (FPP) models is the classifier for face recognition within each individual face class. The bottom face images are the training images in the pose class (the left half profile pose class). Note that the bold lines indicate the face recognition process across pose. . . . . 31

9 Pose classification results using the cosine (Cosine), the $L_1$ (L1), and the $L_2$ (L2) similarity measure, where the means of the five pose classes are derived from the training data. . . . . 32

10 Pose-angled face recognition performance of the Gabor-based kernel PCA method with four polynomial kernels of degree 1 (Md_Gabor_1), degree 2 (Md_Gabor_2), degree 3 (Md_Gabor_3), and degree 4 (Md_Gabor_4), respectively, and the Gabor-based kernel PCA method with two fractional power polynomial models of degree 0.6 (Md_Gabor_0.6) and degree 0.7 (Md_Gabor_0.7), respectively, using the Mahalanobis distance measure. . . . . 33

**Chengjun Liu** received the Ph.D. from George Mason University in 1999, and he is presently an Assistant Professor of Computer Science at New Jersey Institute of Technology. His research interests are in Computer Vision, Pattern Recognition, Image Processing, Evolutionary Computation, and Neural Computation. His recent research has been concerned with the development of novel and robust methods for image/video retrieval and object detection, tracking and recognition based upon statistical and machine learning concepts. The class of new methods he has developed includes the Bayesian Discriminating Features method (BDF), the Probabilistic Reasoning Models (PRM), the Enhanced Fisher Models (EFM), the Enhanced Independent Component Analysis (EICA), the Shape and Texture-based Fisher method (STF), the Gabor-Fisher Classifier (GFC), and the Independent Gabor Features (IGF) method. He has also pursued the development of novel evolutionary methods leading to the development of the Evolutionary Pursuit (EP) method for pattern recognition in general, and face recognition in particular. He is a member of the IEEE and the IEEE Computer Society.