Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Application of Bidirectional Two-dimensional Principal Component Analysis to Curvelet Feature Based Face Recognition

Abdul A. Mohammed, Q. M. Jonathan Wu, Maher A. Sid-Ahmed
Department of Electrical and Computer Engineering
Windsor, Ontario, Canada
{mohammea, jwu, ahmed}@uwindsor.ca

***Abstract* - A bidirectional two-dimensional principal component analysis (2DPCA) is proposed for human face recognition using curvelet feature subspace. Traditionally multiresolution analysis tools namely wavelets and curvelets have been used in the past for extracting and analyzing still images for recognition and classification tasks. Curvelet transform has gained significant popularity over wavelet based techniques due to its improved directional and edge representation capability. In the past features extracted from curvelet subbands were dimensionally reduced using linear principal component analysis (PCA) for obtaining a representative feature set. The novelty of the proposed method lies in the application of 2DPCA to curvelet feature subspace by computing *image covariance matrices* of square training sample matrices in their original form and transposed form respectively to generate a more meaningful and enhanced feature vectors. Extensive experiments were performed using the proposed bidirectional 2DPCA based face recognition algorithm and superior performance is obtained in comparison with state of the art techniques.**

***Keywords*—Principal component analysis, multi-resolution tools, AdaBoost, discrete curvelet transform.**

I. INTRODUCTION

Human face recognition has attracted considerable attention during the last few decades. Human faces represent one of the most common visual patterns in our environment, and humans have a remarkable ability to recognize faces. Face recognition has received significant consideration and is evident by the emergence of international face recognition conferences, protocols and commercially available products. Some of the reasons for this trend are wide range of commercial and law enforcement applications and availability of feasible techniques after decades of research. Typical applications of a face recognition system include driver's license, passports, voter registration card, human-computer interaction, database security, video surveillance; shop lifting, suspect tracking and investigation etc.

Developing a consistent face recognition model is relatively difficult since faces are complex, multidimensional structures and provide a good example of a class of natural objects that do not lend themselves to simple geometric interpretations, and yet the human visual cortex does an excellent job in efficiently discriminating and recognizing these images. Automatic face recognition systems can be classified into two categories namely, constituent and face based recognition [1,2,6]. In the constituent based approach, recognition is achieved based on the relationship between human facial features such as eyes, nose, mouth and facial boundary [5]. The success of this approach relies significantly on the accuracy of the facial feature detection. Extracting facial features accurately is extremely difficult since human faces have similar facial features with subtle changes that make them different from one another.

Face based approaches [4,7] capture and define the image as a whole. The human face is treated as a two-dimensional intensity variation pattern. In this approach recognition is performed through identification and matching of statistical properties. Principal component analysis (PCA) has been proven to be an effective face based approach [3,7]. Kirby *et al.* [7] proposed using Karhunen-Loeve (KL) transform to represent human faces using a linear combination of weighted eigenvectors. Standard PCA based techniques suffer from poor discriminatory power and high computational load. In order to eliminate the inherent limitations of standard PCA based systems, face recognition approaches based on multiresolution tools have emerged and have significantly improved recognition accuracy with a considerable reduction in computation.

Wavelet based approach using PCA for human face recognition [19] proposed by Feng *et al.* utilized a midrange frequency subband for PCA representation and achieved improved accuracy and class separability. In their recent work, Mandal *et al.* [20] has shown that a new multiresolution tool, curvelet along with PCA can be used for human face recognition with superior performance than the standard wavelet subband decomposition. Curvelet transform has better directional and decomposition capabilities than wavelets and has been successfully used for compression and denoising problems. More recently researchers have coined a new technique, namely, two-dimensional principal component analysis (2DPCA) [9] for image representation. As opposed to PCA, 2DPCA is based on 2D image matrices rather than 1D vector so the image matrix does not need to be transformed into a vector prior to feature extraction. Instead, an image covariance matrix is constructed directly using the original image matrices and its eigenvectors are derived for image feature extraction.

Figure 1. Sample images of a subject from FERET database.

In this paper we propose to use coarse level curvelet coefficients together with an application of multi dimensional principal component analysis on the original and transposed covariance matrix for face recognition. Experimental results on five well known face database demonstrate that dimensionally reduced curvelet coefficients using the proposed method offers better recognition in comparison with other PCA based face recognition systems. Fig. 1 shows sample images of a subject from the FERET [24] database; subjects were imaged at different sessions with diverse facial expressions and head rotations.

The remainder of the paper is divided into 5 sections. Section 2 discusses curvelet transform, its variants along with their implementation details followed by a discussion of 2DPCA in section 3. Boosting algorithm for classification is discussed in section 4 and the proposed method is described in section 5. Experimental results are discussed in section 6 followed by conclusion, acknowledgment and references.

II. Feature Extraction with Curvelet Transform

Fourier series decomposes a periodic function into a sum of simple oscillating functions, namely *sines* and *cosines*. In a Fourier series sparsity is destroyed due to discontinuities (Gibbs Phenomenon) and it requires a large number of terms to reconstruct a discontinuity precisely. Multiresolution analysis tools were developed to overcome limitations of Fourier series. Many fields of contemporary science and technology benefit from multiscale, multiresolution analysis tools for maximum throughput, efficient resource utilization and accurate computations. Multiresolution tools render robust behavior to study information content of images and signals in the presence of noise and uncertainty.

Wavelet transform is a well known multiresolution analysis tool capable of conveying accurate temporal and spatial information. Wavelet transform has been profusely used to address problems in data compression, pattern recognition and computer vision. Wavelets better represent objects with point singularities in 1D and 2D space but fail to deal with singularities along curves in 2D. Discontinuities in 2D are spatially distributed which leads to extensive interaction between discontinuities and many terms of wavelet expansion. Therefore wavelet representation does not offer sufficient sparseness for image analysis. Following wavelets, research community has witnessed intense efforts for development of better directional and decomposition tools, namely, contourlets [11] and ridgelets [12]. Curvelet transform [13] is a recent addition to the family of multiresolution analysis tool that is designed and targeted to represent smooth objects with discontinuity along a general curve. Curvelet transform overcomes limitations of existing multiresolution analysis schemes and offers improved directional capacity to represent edges and other singularities along curves. Curvelet transform is a multiscale non-standard pyramid transform with numerous directions and positions at each length and scale. Curvelets outperform wavelets in situations that require optimal sparse representation of objects with edges, representation of wave propagators, image reconstruction with missing data etc. Curvelets have useful geometric features that set them apart from wavelets [13].

A. Continous Time Curvelet Transform

Since the introduction of curvelet transform researchers have developed numerous algorithmic strategies [14-17] for its implementation based on its original architecture. Let us consider a 2D space, i.e. $\Re^2$, with a spatial variable x and a frequency-domain variable ω, and let r and θ represent polar coordinates in frequency-domain. *W(r)* and *V(t)* are radial and angular window respectively. Both windows are smooth, nonnegative, real valued and supported by arguments $r \in [1/2,\ 2]$ and $t \in [-1,\ 1]$. For $j \geq j_0$, frequency window U_j in Fourier domain is defined as,

$$U_j(r,\theta) = 2^{-3j/4}\, W\left(2^{-j} r\right) V\left(\frac{2^{\lfloor j/2 \rfloor}.\theta}{2\pi}\right), \tag{1}$$

where $\lfloor j/2 \rfloor$ is the integral part of $j/2$. Thus the support of U_j is a polar wedge defined by the support of W and V applied with scale-dependent window widths in each direction. Windows W and V always obey the admissibility conditions as follows:

$$\sum_{j=-\infty}^{\infty} W^2\left(2^j\, r\right) = 1, \quad r \in \ (3/4, 3/2)\ . \tag{2}$$

$$\sum_{l=-\infty}^{\infty} V^2(t-l) \ = \ 1, \quad t \in \ (-1/2, 1/2)\ . \tag{3}$$

We define curvelets (as function of $x = (x_1, x_2)$) at a scale 2^{-j}, orientation θ_l, and position $x_k^{(j,l)} = R_{\theta l}^{-1}\left(k_1.2^{-j}, k_2.2^{-j/2}\right)$ by $\varphi_{j,k,l}(x) = \varphi_j\left(R_{\theta_l}\left(x - x_k^{(j,l)}\right)\right)$, where R_θ is an orthogonal rotation matrix. A curvelet coefficient is simply computed by computing the inner product of an element $f \in L^2\left(R^2\right)$ and a curvelet $\varphi_{j,k,l}$,

$$c(j,k,l) = \left\langle f, \varphi_{j,k,l} \right\rangle = \int_{R^2} f(x)\ \overline{\varphi_{j,k,l}}\ dx\ . \tag{4}$$

Curvelet transform also contains coarse scale elements similar to wavelet theory. For $k_1, k_2 \in Z$, we define a coarse level curvelet as:

$$\varphi_{j_0,k}(x) = \varphi_{j_0}\left(x - 2^{-j_0} k\right), \quad \hat{\varphi}_{j_0}(\omega) = 2^{-j_0} W_0\left(2^{-j_0}|\omega|\right). \tag{5}$$

Curvelet transform is composed of fine-level directional elements $(\varphi_{j,k,l})_{j \geq j_0, l, k}$ and coarse-scale isotropic father wavelet $\left(\phi_{j_0,k}\right)_k$. In Fourier space, curvelets are supported near a parabolic wedge. Fig. 2 summarizes the key components of the construction. Shaded area in left portion of Fig. 2 represents a generic wedge. Image on the left represents the induced tiling of the frequency plane and the image on the right shows the spatial Cartesian grid associated with a given scale and orientation. Plancherel's theorem is applied to express $c(j,k,l)$ as integral over the frequency plane as:

$$c(j,k,l) = \frac{1}{(2\pi)^2} \int \hat{f}(\omega) \overline{\hat{\varphi}_{j,k,l}(\omega)}\, d\omega = \frac{1}{(2\pi)^2} \int \hat{f}(\omega) U_j\left(R_{\theta_l}\omega\right) e^{i\left\langle x_k^{(j,l)}, \omega\right\rangle}\, d\omega. \tag{6}$$

B. Fast Discrete Curvelet Transform

Two new algorithms have been proposed in [13] to improve previous implementations. New implementations of FDCT are ideal for deployment in large-scale scientific applications due to lower computational complexity and an utmost 10 fold savings as compared to FFT operating on a similar sized data. We used FDCT via wrapping, described below, in our proposed scheme.

- Apply 2D FFT and obtain Fourier samples $\hat{f}[n_1, n_2]$, $-n/2 \leq n_1,\ n_2 < n/2$.
- For each scale j and angle ℓ , form the product $\tilde{U}_{j,\ell}[n_1, n_2]\hat{f}[n_1, n_2]$.
- Wrap this product around the origin and obtain $\tilde{f}_{j,\ell}[n_1, n_2] = W(\tilde{U}_{j,\ell}\hat{f})[n_1, n_2]$,

  where the range of n_1 and n_2 is $0 \leq n_1 < L_{1,j}$ and $0 \leq n_2 < L_{2,j}$ (for θ in the range of $(-\pi/4, \pi/4)$).
- Apply inverse 2D FFT to each $\tilde{f}_{j,\ell}$, hence collecting the discrete coefficients.

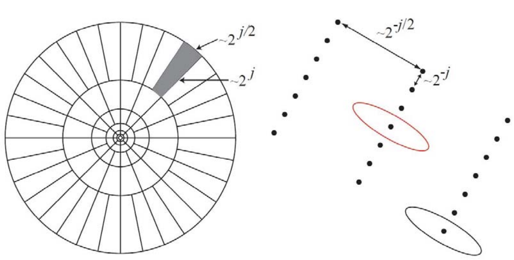

Figure 2. Curvelet tiling of space and frequency [13]

In this work curvelet based features of human faces are extracted using FDCT via the wrapping technique. Coarse level i.e. approximate coefficients are selected for face representation and their dimensionality is reduced using application of 2DPCA on the original and transposed covariance matrix. Application of 2DPCA on the original and transposed covariance matrix is analogous to application of 2DPCA on the original covariance matrix along its rows and columns respectively. Approximate coefficients are selected since they contain an overall structure of the image instead of high frequency detailed information which are insignificant and do not greatly contribute to recognition accuracy.

III. Two Dimensional Princiapal Component Analysis

Karhunen-Loeve expansion, also known as principal component analysis (PCA), is a data representation technique widely used in pattern recognition and compression schemes. Pioneering work by Sirovich and Kirby [7] used PCA for enhanced representation of face images. However PCA cannot capture even a simple invariance unless it is explicitly accounted in the training data. Wiskott *et al.* [8] proposed a *bunch graph matching* technique to overcome limitations and weakness of linear PCA. In [9] Yang *et al.* proposed two dimensional PCA for image representation. As opposed to PCA, 2DPCA is based on 2D image matrices rather than 1D vector so the image matrix does not need to be vectorized prior to feature extraction. Instead an image covariance matrix is computed directly using the original image matrices.

Let X denote a n dimensional unitary column vector. To project an $m \times n$ image matrix A to X ; linear transformation $Y = AX$ is used which results in a m dimensional projected vector Y . The total scatter of the projected samples is introduced to measure the discriminatory power of the projection vector X . The total scatter of the projected samples is characterized by the trace of the covariance matrix of the projected feature vectors i.e. $j(X) = tr(S_x)$, where $tr(\ )$ represents the trace of S_x, and S_x denotes the covariance matrix of the projected feature vectors [13]. The covariance matrix S_x and its trace are computed as:

$$\begin{aligned} S_x &= E\left(Y - E(Y)\right)\ \left(Y - E(Y)\right)^T \\ &= E\left[(A - EA)X\right]\ \left[(A - EA)X\right]^T. \end{aligned} \tag{7}$$

$$tr(S_x) = X^T \left[E(A - EA)^T (A - EA)\right] X. \tag{8}$$

We define $G_t = E\left[(A - EA)^T (A - EA)\right]$ as an image *covariance matrix*. It is easy to verify that G_t is a $n \times n$ nonnegative definite matrix. If there are M training image samples, the j^{th} image sample is denoted by $m \times n$ matrix A_j (where j={1,2,…,M}). Therefore G_t is evaluated according to the following equation.

$$G_t = \frac{1}{M}\sum_{j=1}^{M}\left(A_j - \overline{A}\right)^T\left(A_j - \overline{A}\right) \cdot \quad (9)$$

$$J(X) = X^T G_t X. \quad (10)$$

Where $\overline{A}$ represents the *average image* of all training samples. Above criterion is called the *generalized total scatter criterion.* The unitary vector X that maximizes the criterion is called the optimal projection axis. We usually need to select a set of projection axes, $X_1, X_2, ...X_d$, subject to orthonormal constraint and to maximize the criterion $J(X)$. Yang *et al.* [9] showed that the extraction of image features is computationally more efficient and better recognition accuracy is achieved using 2DPCA than traditional PCA. However the main disadvantage of 2DPCA based recognition is the processing of higher number of coefficients since it works along row directions only. Zhang and Zhou [10] proposed $(2D)^2$ PCA based on assumption that training sample images are zero mean and *image covariance matrix* can be computed from the outer product of row/column vectors of images. In this paper we propose a modified scheme to extract features using 2DPCA by computing two *image covariance matrices* of square training sample matrices in their original form and transposed form respectively, as per the adapted method training image *mean* need not be essentially zero. The vectorization of mutual product of such *covariance matrices* results into a considerably smaller sized feature vector that retains better structural and correlation information amongst neighboring pixels.

IV. AdaBoost Classification

AdaBoost algorithm is an adaptive supervised learning framework which has been successfully implemented for various pattern recognition, computer vision and image processing problems. AdaBoost facilitates powerful incremental learning approach for classification. AdaBoost represents an *ensemble* learning approach formed by a collection of weak learners trained in an iterative fashion where each weak learner is selected based on its classification accuracy on the training set. AdaBoost removes inherent problems in supervised learning such as overtraining, higher error rate, and computational cost; and tenders more emphasis on data that is hard to classify. The main idea of boosting is to combine several weak learners to form an *ensemble* where each weak learner performs slightly better than a random guess. An *ensemble* is formed in a fashion that the performance of individual ensemble member is improved i.e. *boosted.* Suppose we have a set of hypotheses $h_1, h_2, ...h_n$; combined *ensemble* hypothesis takes the form

$$h_f(x) = \sum_{i=1}^{n} \varphi_i h_i(x) \, , \quad (11)$$

where φ_i denotes the weight of each individual *ensemble* member. The idea of boosting finds its roots back to PAC learning algorithm [22]. Main steps involved in AdaBoost method are presented in Table I.

TABLE I. Steps Involved in AdaBoost Algorithm

INPUT: Sequence of N labeled training examples $\{(r_1,s_1),(r_2,s_2),\ldots,(r_N,s_N)\}$ where s_i represents label of example r_i.

- Distribution *D* over the *N* training images
- Weak learning algorithm **WeakLearn**
- Integer T representing maximum number of iterations
- **Initialize** the weight vector: $\varphi_{1,i}=D(i)$ for $i=1,\ldots,N$.
- **Do for** $t=1,2,\ldots,T$

1. $$p_{t,i} = \frac{\phi_{t,i}}{\sum_{i=1}^{N}\phi_{t,i}}$$
2. Call **WeakLearn,** providing it with the distribution $p_{t,i}$; get back a hypothesis $h_{t,i}(r) = R \rightarrow [0,1]$
3. Calculate the error of $h_{t,i}$:
4. $$\varepsilon_{t,i} = \sum p_{t,i}\left|h_{t,i}(r) - s_i\right|$$
5. $$\beta_{t,i} = \frac{\varepsilon_{t,i}}{1-\varepsilon_{t,i}}$$
6. Set the new eight vectors to be
7. $$\phi_{t+1,i} = \phi_{t,i}\beta_{t,i}^{1-\left|h_{t,i}(r)-s_i\right|}$$

OUTPUT: Classifier - hf(x)

$$h_f(r) = \begin{cases} 1 & \text{if } \sum_{t=1}^{T}(\log 1/\beta_t) h_t(r) \geq \frac{1}{2}\sum_{t=1}^{T}\log 1/\beta_t \\ 0 & \text{otherwise} \end{cases}$$

V. Proposed Method

Our proposed method deals with classification of face images using AdaBoost scheme utilizing dimensionally reduced feature vectors obtained from curvelet space. Images from each database are converted into gray level image with 256 gray levels. Conversion from RGB to gray level format was the only pre-processing performed on the images. In addition to the mentioned alteration there were no further changes made to the images as it may lead to image degradation. We randomly divide image database into two sets namely training set and testing set. Recently, research community has observed dimensionality reduction techniques being applied on data to be classified for real-time, accurate and efficient processing. All images within each database have the same dimension, i.e. *R*x*C*. Similar image sizes support the assembly of equal sized curvelet coefficients and feature vector extraction with identical level of global content. Curvelet transform of every image is computed and only coarse level coefficients are extracted. Curvelet transform is a relatively new technique for multiresolution analysis that better deals with singularities in higher dimension, and enhances localization of higher frequency components with minimized aliasing effects.

Application of AdaBoost classification algorithm on original curvelet vectors could be computationally expensive due to higher dimensionality of data originating from large image database. Outliers and irrelevant image points being included into classification task can also affect the performance of our algorithm; hence 2DPCA is employed to reduce dimensionality of curvelet vectors. 2DPCA was proposed in the pioneering work of [9] wherein an image covariance matrix is computed directly using the original image matrices. Features are extracted by computing two *image covariance matrices* of square training sample matrices in their original and transposed form respectively; mutual product of such *covariance matrices* retains better structural and correlation information amongst neighboring pixels. Dimensionally reduced curvelet coefficients are vectorized into an UxV dimension vector, final feature vector, where Ux$V << R$xC.

2DPCA based feature vectors better retain the *global* structure of input space and facilitate accurate classification with lower computational complexity, diminished outliers and irrelevant information. An AdaBoost classification algorithm is trained using labeled 2DPCA feature vectors computed in step 4 (Table II). Table II consists of detailed steps that demonstrate our proposed technique.

VI. EXPERIMENTAL RESULTS

Extensive experiments were performed using our proposed method on 5 distinctive face database namely, FERET, AT&T [25], Georgia Tech [26], Faces94 [27] and JAFFE [28] data sets. These database are some of the popular database used in literature to evaluate and compare face recognition algorithms with existing state of the art algorithms. Before divulging into experimental details and results achieved using the proposed method, we will briefly describe the database used to vigorously test our algorithm.

TABLE II. STEPS INVOLVED IN THE PROPOSED ALGORITHM

INPUT: Randomly divide image database into two subsets TR_i and TE_j where i={1,2,…,n} and j={1,2,…,m} representing training and test image, each of size RxC, sets respectively.

OUTPUT: Classifier - f(x)

1. Compute the curvelet transform of each training and test images and extract coarse level feature sets. Each feature set is of dimension UxV << RxC.

   (find detailed discussion of curvelet transform in section 2)

2. Calculate image covariance matrix of test and train images in their original and transposed form.

   $$G_{tR} = \frac{1}{n}\sum_{i=1}^{n}(A_i - \overline{A_R})^T(A_i - \overline{A_R})$$

   $$G_{tE} = \frac{1}{m}\sum_{j=1}^{m}(A_j - \overline{A_E})^T(A_j - \overline{A_E})$$

3. Evaluate the maximizing criteria *J(X)* for train and test images.

   $$J(X_R) = X_R^T G_{tR} X_R$$

   $$J(X_E) = X_E^T G_{tE} X_E$$

4. Obtain 2DPCA based feature vectors, *fp*, by computing principal components of every original image covariance matrix and transposed image covariance matrix respectively.
5. Train AdaBoost classifier: Generate set of positive and negative samples of 2DPCA based feature vectors (vectorized feature vectors obtained in previous step) for training.
6. Classify images with test feature vectors using AdaBoost trained in step 5.

A. Database

The FERET database was sponsored by the Department of Defense in order to develop a system with automatic face recognition capability to be employed for assistance in security, intelligence and law enforcement. The final corpus consists of 14051 eight-bit grayscale images of human heads with views ranging from frontal to left and right profiles, see [24] for more details.

AT&T face database contains 10 different images for each of the 40 distinctive subjects. Images of some subjects were taken at different times, with varying lighting conditions, facial expressions and facial details. All images were taken against a dark homogeneous background with the subjects in an upright, frontal position with a small tolerance for side movement.

Georgia Tech database contains images of 50 people and contains 15 color images for every subject. Most of the images were taken in two different sessions to take into account the variations in illumination conditions, facial expression, and appearance. In addition to this, images were captured at varying scales and orientations.

Faces94 database was generated at University of Essex and contain a series of 20 images per individual. Faces94 database is wide-ranging and contains 20 images of 152 distinctive individuals. The database contains images of people of various racial origins, mainly first year undergraduate students, so the majority of individuals are between 18-20 years old but some older staff member and students are also present. Some individuals are wearing glasses and/or beards.

Finally a Japanese female facial expression (JAFFE) database is also used to rigorously test the performance of the proposed method. The database contained 220 images of varying facial expressions posed by 10 Japanese female models.

B. Quantitative results and discussion

All Images are converted from RGB to gray level in our experiments. In the FERET database different number of images exists for different subjects so 30% of images from each subject were used as prototypes and the remaining 70% for testing. Similarly 30% of images of each subject from the AT&T, Faces94, Georgia Tech and JAFFE database were used for training and the remaining 70% are used for testing the accuracy of our proposed face recognition algorithm. Both the testing and training sets of images are decomposed using curvelet transform at 3 scales and 8 different angles. Amongst the curvelet coefficients only approximate coefficients are selected as initial feature vectors since they closely represent and approximate the input image. The selected feature vectors are dimensionally reduced through application of 2DPCA along image rows and columns respectively (application of 2DPCA on original and transposed image covariance matrix). Dimensionally reduced feature vectors are vectorized and AdaBoost classification is performed. The above process was repeated 4-5 times for all database and averaged results included. The recognition accuracy for Faces94 and JAFFE database using our proposed method is compared with curvelet based linear PCA, wavelet based linear PCA and linear PCA, listed in Table III-IV.

TABLE III. RECOGNITION ACCURACY FOR JAFFE DATABASE

Number of Principal Components	Average Recognition Rate (%)			
	PCA [7]	*Wavelet + PCA [19]*	*Curvelet + PCA [20]*	*Curvelet + 2DPCA+ AdaBoost (Proposed)*
5	92.29	98.67	91.71	94
10	96.18	99.21	96.92	97.34
15	96.19	99.38	98.71	98.66
25	98.31	99.57	99.49	100
30	98.12	99.6	99.74	100
40	98.26	99.6	100	100
50	98.26	99.6	100	100
70	98.26	99.6	100	100
90	98.26	99.6	100	100

TABLE IV. RECOGNITION ACCURACY FOR FACES94 DATABASE

Number of Principal Components	Average Recognition Rate (%)			
	PCA [7]	*Wavelet + PCA [19]*	*Curvelet + PCA[20]*	*Curvelet + 2DPCA+ AdaBoost (Proposed)*
5	89.87	93	98.43	99.12
10	90	94	99.14	99.39
15	93.9	98	99.16	99.53
25	96.84	99	99.23	100
30	98	99.23	99.25	100
40	98	99.25	99.28	100
50	98	99.25	99.27	100
70	98	99.25	99.30	100
90	98	99.26	99.30	100

Curvelet based linear PCA is implemented by decomposing test and train images using curvelet transform at 3 scales and 8 different angles. Coarse level curvelet coefficients are selected, vectorized and dimensionally reduced using linear PCA. Similarly wavelet based linear PCA is implemented using a multi level wavelet subband decomposition of images and only the lower frequency subimage is used for dimensionality reduction using linear PCA. In linear PCA, images are not transformed to a different domain and principal eigen vectors of face images are used as features for recognition. Varying number of principal components are used to emphasize the recognition accuracy achieved using two-dimensional PCA and linear one-dimensional PCA based techniques prior to saturation. It is clearly evident that the accuracy of our proposed method is better than that achieved using linear PCA based techniques. We can conclude from the above results that curvelet based linear PCA face recognition algorithm performs better than wavelet based and standard one-dimensional PCA procedure, although slight inferior than our proposed method, therefore we compare our proposed method with curvelet based one-dimensional linear PCA for the remaining database to establish superiority of our proposed system.

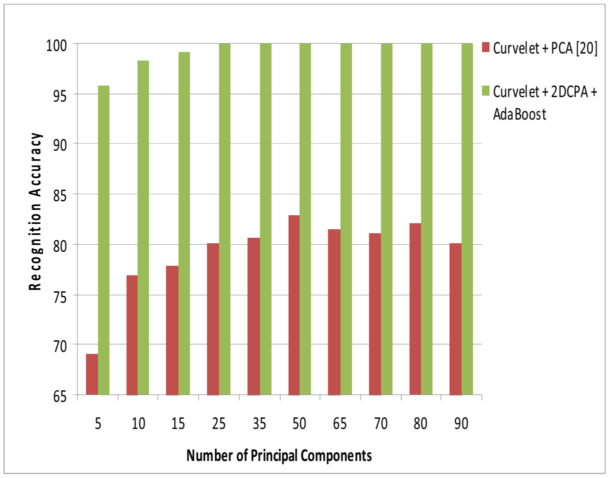

Figure 3. Recognition accuracy for FERET database

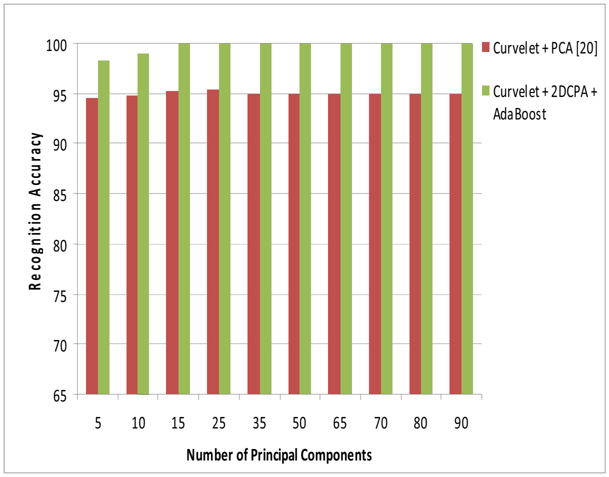

Figure 4. Recognition accuracy for Georgia tech database

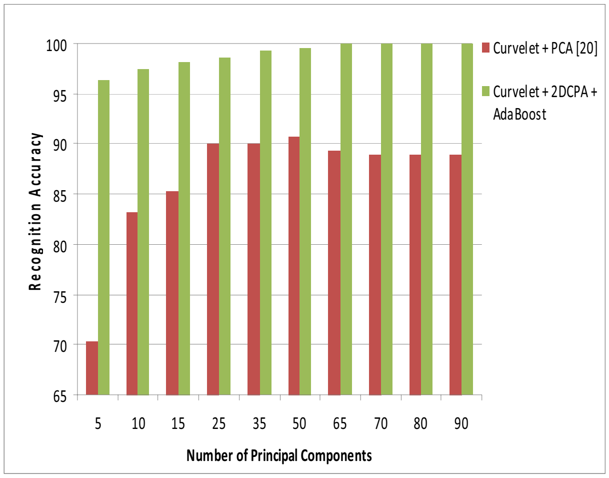

Figure 5. Recognition accuracy for AT&T database

Figs. 3-5 show recognition accuracy obtained using curvelet based PCA and our proposed method is compared for FERET, Georgia Tech. and AT&T database respectively. From the above results it is evident that our proposed method for face recognition performs radically better recognition in comparison with a curvelet based linear PCA face recognition system. Our proposed method achieves improved recognition accuracy with considerably smaller number of principal components when compared to other state of the art recognition algorithms, thereby achieving a significant savings in computational cost during classification at a required recognition rate. Improvements in recognition rate are significantly attributed to the use of two-dimensional PCA that treats the curvelet coefficient matrix as a single unit instead of converting it into a series of one dimensional vectors and treating them independently. A two-dimensional PCA compactly retains relationship amongst curvelet coefficients and generates an enhanced representative feature set for classification.

VII. Discussion

We propose a novel face recognition technique using nonlinear curvelet feature subspace. Curvelet transform is used as multiresolution analysis tool to compute sparse features. Localized high frequency response with minimized aliasing, better directionality, and improved processing of singularities along curves demonstrate the superior performance of curvelet transform as feature extractor. Two-dimensional PCA is utilized for dimension reduction and applied to original image covariance matrix and its transposed version to generate an accurate representative feature set. AdaBoost classification scheme is employed for ascertaining recognition/classification and to eliminate inherent problems that arise in supervised learning such as overtraining, higher error rate and computational cost. Experiments are performed using five popular human face database and significant improvement in recognition accuracy is achieved. The proposed method drastically outperforms conventional face recognition systems that employ linear one-dimensional PCA. Law enforcement, intelligence and security agencies can potentially benefit from our proposed recognition scheme.

Acknowledgment

This research has been supported in part by the Canada Research Chair Program and the Natural Sciences and Engineering Research Council of Canada discovery grant. The authors would also like to extend their gratitude to curvelet.org team for helpful data and links.

References

[1] B. S. Manjunath, R. Chellappa and C. V. Malsburg, “A feature based approach to face recognition,” Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 1992, pp. 373-378.

[2] G. Chow and X. Li, “Towards a system for automatic facial feature detection,” Pattern Recognition 26(12), 1993, pp. 1739-1755.

[3] A. Pentland, B. Moghaddam and T. Starner, “View-based and modular eigenspaces for face recognition,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, Seattle, 1994, pp. 84-91.

[4] D. L. Swets and D. L. Weng, “Using discriminant eigenfeatures for image retrieval,” IEEE Trans. Pattern Anal. Mach. Intell. 18(8), 1996, pp. 831-836.

[5] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma and N. Otsu, “Face recognition system using local autocorrelations and multiscale integration,” IEEE Trans. Pattern Anal. Mach. Intell. 18(10), 1996, pp. 1024-1028.

[6] W. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips, “Face recognition: a literature survey,” ACM Computing Surveys, 2003, pp. 399-458.

[7] M. Kirby and L. Sirovich, “Application of the Karhunen-Loeve procedure for the characterization of human faces,” IEEE Trans. Pattern Anal. Mach. Intell. 12, 1990, pp. 103-108.

[8] L. Wiskott, J.M. Fellus, N. Kruger, and C. Von der Malsburg, “Face Recognition by Elastic Bunch Graph Matching,” IEEE Trans. on PAMI, Vol. 19, no.7, 1997, pp. 775-779.

[9] J. Yang, D. Zhang, A. F. Frangi and J.Yang, “Two-dimensional PCA: a new approach to appearance based face representation and recognition,” IEEE Trans. Pattern Anal. Mach. Intell. 26(1), 2004, pp. 131-137.

[10] D. Zhang, Z. H. Zhou, “$(2D)^2$ PCA: Two-directional two-dimensional PCA for efficient face representation and recognition,” Neurocomputing, Vol. 69, No. 1-3, 2005, pp. 224-231.

[11] M. N. Do and M. Vetterli, “The contourlet transform: An efficient directional multiresolution image representation,” IEEE trans. on image processing, 14(12), 2005, pp. 2091-2106.

[12] M. N. Do and M. Vetterli, “The finite ridgelet transform for image representation,” IEEE trans on image processing. 12(1), 2003, pp. 16-28.

[13] E. J. Candès, L. Demanet, D. L. Donoho and L. Ying, “Fast discrete curvelet transforms,” Multiscale Model. Simul.,2005, pp. 861-899.

[14] J. L. Starck, N. Aghanim and O. Forni, “Detecting cosmological non-Gaussian signatures by multi-scale methods,” Astronomy and Astrophysics 416(1), 2004, pp. 9-17.

[15] J. L. Starck, M. Elad and D. L. Donoho, “Redundant multiscale transforms and their application for morphological component analysis,” Advances in Imaging and Electron Physics 132, 2004, pp. 287-342.

[16] F. J. Herrmann and E. Verschuur, “Separation of primaries and multiples by non-linear estimation in the curvelet domain,” EAGE 66th Conference & Exhibition Proceedings,2004.

[17] H. Douma and M. V. Hoop, “Wave-character preserving prestack map migration using curvelets,” Presentation at the Society of Exploration Geophysicists, Denver, 2004.

[18] T. M. Cover, “Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition,” IEEE Transaction on Electronic Computers, 14(3), 1965, pp. 326-334.

[19] G. C. Feng, P. C. Yuen and D. Q. Dai, “Human face recognition using PCA on wavelet subband,” Journal of Electronic Imaging, 9(2), 2000, pp. 226-233.

[20] T. Mandal and Q. M. J. Wu., “Face Recognition using Curvelet Based PCA,” ICPR, Tampa, 2008.

[21] Y. Freud and R. E. Schapire, “A decision theoretic generalization of on-line learning and an application to boosting,” Journal of Computer and system sciences 55, 1997, 119-139.

[22] L. G. Valiant, “A theory of learnable,” Artificial intelligence and language processing, Vol 27, 1984, 134-142.

[23] S. Ali and M. Shah, “A supervised learning framework for generic object detection in images,” International conference on Computer Vision, China, Oct. 2005.

[24] http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

[25] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[26] http://www.anefian.com/face_reco.htm

[27] http://cswww.essex.ac.uk/mv/allfaces/index.html

[28] http://www.kasrl.org/jaffe.html