



## Curvelet based face recognition via dimension reduction

Tanaya Mandal<sup>a,\*</sup>, Q.M. Jonathan Wu<sup>b</sup>, Yuan Yuan<sup>c</sup>

<sup>a</sup> ECE, University of British Columbia, Kaiser 2010, 2332 Main Mall, BC, Canada

<sup>b</sup> ECE, University of Windsor, ON, Canada

<sup>c</sup> School of Engineering and Applied Science, Aston University, Birmingham, UK

### ARTICLE INFO

#### Article history:

Received 31 October 2008

Received in revised form

9 March 2009

Accepted 9 March 2009

Available online 18 March 2009

#### Keywords:

Digital curvelet transform

LDA

Multiresolution analysis

PCA

Subbands

Wavelet transform

### ABSTRACT

Multiresolution ideas, notably the wavelet transform, have been proved quite useful for analyzing the information content of facial images. Numerous papers and research articles have discussed the application of wavelet transform in face recognition. However, little attention has been paid to the newly developed multiresolution tools (contourlet, curvelet, etc.) despite their improved directional elements and other promising abilities compared to traditional wavelet transform. In this article we introduce the application of digital curvelet transform in conjunction with different dimensionality reduction tools, looking particularly at the problem of facial feature extraction from 2D images. The purpose of this paper is exploratory. We do not claim that the results achieved here are the best possible. Rather, we aim at showing that curvelets can serve as an effective alternative to wavelets as a feature extraction tool. This work can be seen as a stepping stone for further research in this direction. Our methods have been evaluated on well-known databases like ORL, Essex Grimace and Yale face. Curvelet based results have been compared with that achieved using wavelets and other existing techniques to show that curvelets indeed has the potential to supersede wavelet based results.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

Face recognition has been studied diligently for more than 30 years now and has emerged as one of the most successful applications of image analysis. Compared to other biometrics (fingerprint or iris recognition), face recognition might not have a superior level of accuracy, but working with faces certainly has a number of advantages. Firstly, the data (facial images) can be collected using a simple camera (compare the cost and difficulties of collecting good fingerprints or iris scan); and secondly, the images can be collected even without the knowledge of the subject. Clearly, from the security point

of view, face recognition is a means of biometric identification which is hassle-free and inexpensive. The subject has become a major issue in the past decade—due to its important real-world applications in areas like video surveillance, smart cards, database security, telecommunication, digital libraries and medical records [1]. Usually, the available images are 2D intensity images of human faces, which are 3D objects. So this problem can also be seen as a task of identifying 3D objects from their 2D images.

Feature extraction is a key step prior to face recognition. Extraction of a representative feature set can greatly enhance the performance of any face recognition system. Direct use of pixel values as features is not possible due to huge dimensionality of the images. Traditionally, Principal Component Analysis (PCA) is employed to obtain a lower dimensional representation of the data in standard eigenface based methods [2]. Though PCA provides

\* Corresponding author.

E-mail addresses: [tanaya@ece.ubc.ca](mailto:tanaya@ece.ubc.ca) (T. Mandal), [jwu@uwindsor.ca](mailto:jwu@uwindsor.ca) (Q.M. Jonathan Wu), [yuany1@aston.ac.uk](mailto:yuany1@aston.ac.uk) (Y. Yuan).

effective approximation, the method suffers from greater computational load and poor discriminatory power [3]. In order to resolve these limitations of PCA, researchers suggest the use of other dimensionality reduction tools like Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Kernel PCA [4] and Kernel LDA [5]. ICA is a generalization of PCA and finds not-necessarily orthogonal basis to represent the data. Based on the belief that ICA is capable of providing a better probabilistic model it was used in [6] to extract facial features and a performance improvement over PCA was reported. Belhumeur et al. worked with Fisherfaces [7], which involves Fisher Linear Discriminant (FLD). LDA has been found to improve the classification accuracy of a system when multiple images are available per class. LDA based recognition methods aim at simultaneously maximizing between-class scatter and minimizing within-class scatter. KPCA, which takes into account higher order statistics in contrast to PCA, has also been successfully applied in the context of face recognition [8]. More recently, a faster and efficient technique called Binary Two-Dimensional PCA (B-2DPCA) has been proposed [9]; this method relies on Two-Dimensional PCA (2DPCA) based [10] decomposition for fast numerical computations. The algorithm is particularly suitable for large-scale and high-resolution image based biometric systems. The problem of high dimensional feature extraction has been very well-addressed in [11,12]. Most recently tensor based approaches [13,14] have been quite popular. Motivated by the success of 2DLDA [15], General Tensor Discriminant Analysis (GTDA) has been developed in [13]. This method uses an alternating projection optimization algorithm, which, unlike 2DLDA always converges to a solution. In [14], another new manifold learning method called Discriminant Locally Linear Embedding (DLLE) has been presented. In this technique, the intra-class local geometric properties are preserved according to a locally linear embedding criterion and inter-class separability is enforced by maximizing margins between point pairs on different classes [14]. The method has been reported to achieve significant performance improvement compared to state-of-the-art gait recognition algorithms.

Another way to handle huge dimensionality in face recognition problems is to employ dimensionality reduction tools on some kind of transformation domain. In [16], the authors argued that, when raw images are subjected to PCA, the correlation of facial features is not well-reflected in eigenspace. So, they suggested using Gabor filter to extract facial features first and then use PCA to classify the features optimally. It has been claimed that several problems like, deformation of face images due to in-plane in-depth rotation, illumination and contrast variation can be solved by extracting facial features using Gabor filters. Nowadays, multiresolution analysis is often performed as a pre-processing step to dimensionality reduction. The most popular multiresolution analysis tool is the wavelet transform. It has enjoyed a wide-spread popularity in the field of face recognition [17,18]. In wavelet analysis an image is usually decomposed at different scales or resolutions using a wavelet basis vector. Thereafter, the component which corresponds to the maximum variance is selected for

further operation. This way, the image can be represented by a small number of wavelet coefficients and the effect of variable facial appearances (expression variation, illumination variation, facial detail variation, etc.) on the classification systems can also be lessened in turn. Another approach has been introduced of late by Pang et al. in [19]. They have used Gabor-based Region Covariance Matrices (RCM) for face recognition problems. From the experiments performed on FERET, it has been shown that the proposed methodology provides significant improvement over the conventional RCM framework.

Over the past two decades, following the success of wavelets, a series of new multiresolution analysis tools like, ridgelets [20], contourlets [21], etc. were developed. 'Curvelet transform' [22–24], developed by Candes and Donoho is a recent addition to this list. Compared to wavelets, curvelets have improved directional elements and better ability to represent edges and other singularities along curves. Curvelets, being a relatively new concept has not yet been very popular. So far, its successful applications have been found mostly in the fields of image processing, e.g. image denoising [23], image compression [25], image fusion [26], contrast enhancement [27], image deconvolution [28], high quality image restoration [29], astronomical image representation [30] etc; but not much work has been done to explore the potential of curvelet transform to solve pattern recognition problems. In some recent works, Majumdar showed that curvelets can serve as the basis for pattern recognition problems like character recognition [31]. Curvelet based face recognition has been discussed in [32–34]. In [33], a SVM based face recognition system has been developed, which uses curvelet coefficients without employing any dimensionality reduction. In our previous work [32], curvelet transform has been employed to extract features from bit quantized facial images, where we showed that curvelets can indeed supersede the performance of wavelets. The work presented in [34] is an extension of [32]. In [34], the authors used the bit-quantized images to extract curvelet features at five different resolutions. A total of 15 sets of approximate coefficients are collected and used to train 15 Support Vector Machines. The results are fused using majority voting. However, working with large number of curvelet features (as done in all previous face recognition works with curvelets) can be computationally expensive. So, the natural aim is to reduce the dimension further. In this article a thorough study of application of digital curvelet transform in conjunction with dimensionality reduction tools like PCA and LDA will be presented. The proposed methods have been evaluated by carrying out different experiments on three well-known databases: ORL, Essex Grimace and Yale Face Database. Our prime objective has been to explore the potential of curvelet transform to be used as an alternative feature extraction tool.

The rest of the paper is organized as follows: Section 2 discusses the idea of curvelet transform, followed by a comparative study of curvelets and wavelets in Section 3. The proposed methods are described in Section 4. Section 5 contains experimental results and discussion; Section 6 concludes the article and suggests future work.

## 2. Curvelet transform

Curvelet transform was developed by Candes and Donoho in 1999. Its development was motivated by the need of image analysis [22]. The transform has improved directional capability, better ability to represent edges and other singularities along curves as compared to other traditional multiscale transforms, e.g. wavelet transform. In the past few years curvelet construction has been redesigned in order to make it simpler to understand and use. This second generation curvelet transform [35], introduced in 2006 is not only simpler, but is faster and less redundant compared to its first generation version [24]. Curvelet transform is multiscale and multidirectional. Curvelets exhibit highly anisotropic shape obeying parabolic-scaling relationship (they take the shape of elongated needles at finer scales). In order to implement curvelet transform, first a 2D FFT of the image is taken. Then the 2D Fourier frequency plane is divided into ‘parabolic’ wedges. Finally an inverse FFT of each wedge is taken to find the curvelet coefficients at each scale  $j$  and angle  $\ell$ . Fig. 1 (left) shows the division of wedges of the Fourier frequency plane. The wedges are the result of partitioning the Fourier plane in radial (concentric circles) and angular divisions. Concentric circles are responsible for decomposition of the image in multiple scales (used for bandpassing the image) and angular divisions correspond to different angles or orientation. Hence, to address a particular wedge one needs to define the scale and angle first. In the spatial domain, each wedge corresponds to a particular curvelet at that given scale and angle. Fig. 1 (right) represents curvelets in spatial Cartesian grid associated with a given scale and angle [35].

There are two different digital implementations of Fast Digital Curvelet Transform (FDCT) [35]:

- Curvelets via USFFT (Unequally Spaced Fast Fourier Transform).
- Curvelets via Wrapping.

These transforms are linear and take as input a Cartesian array  $f[t_1, t_2]$ ,  $0 \leq t_1, t_2 < n$  and provide an output of discrete coefficients. Though both the implementations use same digital coronization, they differ in the choice of spatial grid to translate curvelets at each scale and angle. Both the FDCTs run in  $O(n^2 \log n)$  flops for  $n$  by  $n$  Cartesian arrays

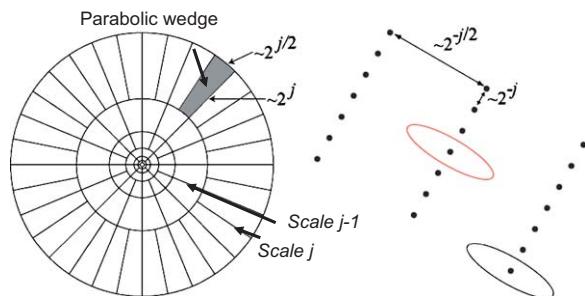


Fig. 1. Curvelets in Fourier frequency (left) and spatial domain (right) [35].

[35]. In case of wrapping a rectangular grid is assumed. All the experimental work presented in this article use numerically tight FDCT Wrapping, as this is the fastest curvelet transform algorithm currently available [35]. The algorithm has been described below.

Let  $\hat{f}[n_1, n_2]$  denote 2D discrete Fourier transform of  $f[t_1, t_2]$ . Let  $U_j(\omega)$  be a localizing window and  $\tilde{U}_j[n_1, n_2]$  is supported on some rectangle of length  $L_{1,j}$  and width  $L_{2,j}$  [35]

$$P_j = \{(n_1, n_2) : n_{1,0} \leq n_1 < n_{1,0} + L_{1,j}, n_{2,0} \leq n_2 < n_{2,0} + L_{2,j}\} \quad (1)$$

Implementation steps of FDCT via Wrapping [35]:

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

$$\hat{f}_{j,\ell}[n_1, n_2] = W(\tilde{U}_{j,\ell} \hat{f})[n_1, n_2], \quad (2)$$

where the range  $n_1$  and  $n_2$  is now  $0 \leq n_1 < L_{1,j}$  and  $0 \leq n_2 < L_{2,j}$ .

- (4) Apply the inverse 2D FFT to each  $\hat{f}_{j,\ell}$ , hence collecting the discrete coefficients.

In first two steps, the Fourier frequency plane of the image is divided into radial and angular wedges owing to the parabolic relationship between curvelet’s length and width, as shown in Fig. 1. Each wedge corresponds to curvelet coefficients at a particular scale and angle. Step 3 is basically re-indexing the data around the origin. Finally, using inverse FFT, discrete curvelet coefficients are collected in the spatial domain. It is not possible to delve into the mathematical details of curvelet transform within the scope of this paper; interested reader may refer to the works of Candes and Donoho [22–24,35].

## 3. Wavelet vs. curvelet

The sparsity of Fourier series is destroyed due to discontinuities (Gibbs phenomenon); it requires a large number of terms to reconstruct a discontinuity in Fourier series within a good accuracy. Later, wavelets are found to have the ability to solve the problems of Fourier series, as they are localized and multiscale. However, though wavelets do work efficiently in one-dimension, they fail to represent higher dimensional singularities effectively due to limited orientation selectivity. Wavelets and related classical multiresolution ideas exploit a limited dictionary made up of roughly isotropic elements occurring at all scales and locations [36]. These dictionaries do not exhibit highly anisotropic elements and there are only a fixed number of directional elements (standard orthogonal wavelet transforms have wavelets with primarily vertical, primarily horizontal and primarily diagonal orientations) independent of scale. Images do not always exhibit isotropic scaling and thus these limitations of wavelets call for other kinds of multiscale representation.

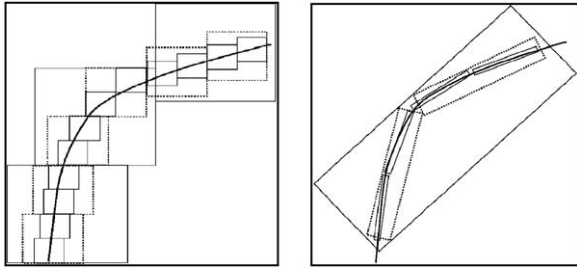


Fig. 2. Edge representation by wavelet and curvelet transform [37].

The most interesting fact about curvelets is that it has been developed specially to represent objects with ‘curve-punctuated smoothness’ [36] i.e. objects which display smoothness except for discontinuity along a general curve; images with edges would be good example of this kind of objects. Wavelet transform has been profusely employed to address different problems of pattern recognition and computer vision because of their capability of detecting singularities. But, though wavelets are good at representing point singularities in both 1D and 2D signals, they fail to detect curved singularities efficiently. Fig. 2 shows the edge representation capability of wavelet (left) and curvelet transform (right). For the square shape of wavelets at each scale, more wavelets are required for an edge representation than that compared to the number of required curvelets, which are of elongated needle shape. One more novelty of curvelet transform is that it is based on anisotropic scaling principal, whereas wavelets rely on isotropic scaling.

Let us consider an image function  $f$ , which has a discontinuity across a curve and is otherwise smooth; if it is approximated by the best  $m$  terms in Fourier expansion, the squared error is given by [37]

$$\|f - \tilde{f}_m^F\|^2 \propto m^{-1/2}, \quad m \rightarrow +\infty$$

For wavelets,

$$\|f - \tilde{f}_m^W\|^2 \propto m^{-1}, \quad m \rightarrow +\infty$$

For curvelets, we have

$$\|f - \tilde{f}_m^C\|^2 \propto \log(m^3)m^{-2}, \quad m \rightarrow +\infty$$

which approaches the ideal reconstruction rate asymptotically. The main idea here is that the edge discontinuity is better approximated by curvelets than wavelets. To summarize, wavelet transform suffers from the following limitations:

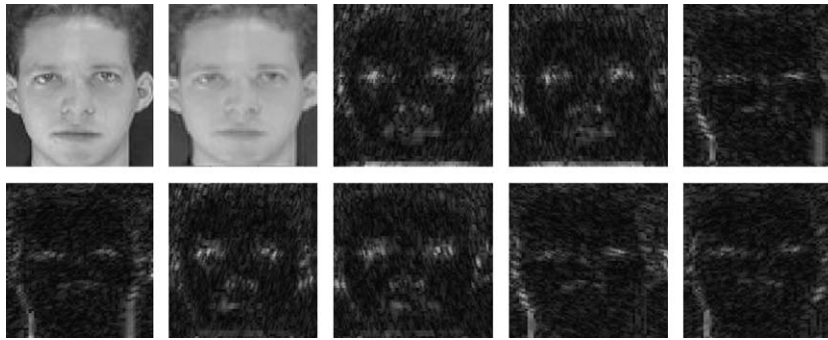
- Edge representation—though wavelets perform better than FFT, it is not optimal.
- Crude directional elements independent of scale.
- No highly anisotropic element.

Curvelet transform is capable of solving the above problems. At any scale  $j$ , curvelets provide a sparse representation  $O(2^{j/2})$  of the images compared to wavelets  $O(2^j)$ . Now, we will empirically try to show that similar (or

better) accuracy can be achieved using curvelet transform as the bases of face recognition.

#### 4. Curvelet based face recognition

In the previous section, we have argued theoretically, that curvelet transform, because of its improved directional elements and highly anisotropic scaling property can provide an equally good (or superior) basis for face recognition compared to wavelets. To validate this empirically, various experiments have been designed and performed. Our face recognition system is divided into two stages: training stage and classification stage. During training, curvelet transform is applied to decompose the images into curvelet subbands. It has been stated repeatedly that curvelets are very good at representing objects with ‘curve-punctuated smoothness’ [36], i.e. objects which display smoothness except for discontinuity along a general curve. Facial images with edges are good examples of this kind of objects. Say, the images are represented by eight bits, i.e. 256 gray levels. In such an image two very close regions can have differing pixel values. Such a gray scale image will then have a lot of ‘curved edges’ (which are typically curved in facial images)—and consequently the curvelet transform will capture this crucial edge information. Extracted curvelet coefficients can then be directly fed to a classifier to perform the identification task. An image of size, say  $64 \times 64$ , when decomposed using curvelet transform at scale 3 (coarse, fine, finest) and angle 8 will produce 1 approximate (coarse) subband of size  $21 \times 21$  and 24 detailed coefficients of slightly larger size. Working with such large number of feature vectors is extremely expensive. Hence it is of utmost importance to determine a representative set of features. Selection of subbands has been done on the basis of the amount of variance they account for. We have selected the coarse coefficients, (as they account for the maximum variance and contain maximum energy of the image-data) and a set of fine coefficients with next highest amount of total variance. Thus selecting the appropriate curveletfaces reduces the dimension of the original image (represented as a vector) from  $1 \times 4096$  to  $1 \times 1258$ . Yet a feature vector of length  $1 \times 1258$  needs some kind of dimensionality reduction. Traditional dimensionality reduction methods like PCA/LDA can be applied on the selected subbands thereafter and an efficient and representative dataset can be produced. A simple Nearest Neighbor (NN) classifier is used to carry out the recognition task. Unless mentioned otherwise, a scale value of 3 has been used through out this work, in order to strike a balance between the speed and performance of the system. It has been observed that higher scale (4 or 5) has shown only marginal improvement or even no improvement in terms of recognition accuracy. Fig. 3 shows the curvelet subbands for a sample face taken from ORL dataset. These curvelet decomposed images are named ‘CURVELETFACES’. Digital curvelet transform (scale = 2, angle = 8) is applied on the original image of size  $112 \times 92$ . This produces 1 approximate ( $75 \times 61$ ) and eight detailed curvelet coefficients (four of



**Fig. 3.** Curvelet transform of faces: first image in the first row is the original image, second image in the first row is the approximate coefficients and others are detailed coefficients at eight angles; all the images are resized to same dimension for the purpose of illustration only.

them are of size  $66 \times 123$  and rest four are of size  $149 \times 54$ ) for eight different angles. Fig. 3 shows these nine curvelet coefficients. The images though are of different size, have been resized to same size only for the sake of neat presentation.

#### 4.1. Recognition using curvelet based PCA

PCA has been successfully applied on wavelet decomposed images which resolves the limitations of using PCA directly [3,38]. Inspired by the success of subband PCA and superiority of curvelets over wavelets, PCA has been applied on curvelet decomposed images for dimensionality reduction; a representational basis is formed. Curvelet transform presents the edge information efficiently and forms a set of distinctive features for the images. In classification stage the test images are subjected to the same operations and they are transformed to the same representational basis. A simple distance based classifier like NN classifier has been employed. Curvelet parameters are set to scale = 3 and angle = 8. PCA has been applied on the selected subbands. This solves the problem of computational load. With PCA, it is always an issue to select the optimum number of principal components. This has been studied by varying the number of eigenvectors and showing the corresponding variation in recognition accuracy.

#### 4.2. Recognition using curvelet based LDA

Researchers argue that PCA suffers from poor discriminatory power and prefer LDA over PCA. LDA is said to be able to improve inter-class separability by maximizing within-class similarity as well as between-class variance. We have employed LDA in stead of PCA this time. After decomposing the images using curvelet transform, LDA has been performed on the resulting approximate curveletfaces. A matrix  $S = S_w^{-1} S_b$  is derived, where  $S_w$  is the within-class scatter matrix and  $S_b$  is the between-class scatter matrix [39]. Then  $n$  numbers of eigenvectors corresponding to largest  $n$  eigenvalues are selected. During the experiments  $n$  has been varied to study the change in recognition accuracy.

#### 4.3. PCA+LDA framework

LDA, though successfully applied for face recognition problems, encounters two major difficulties [40]. Firstly, the size of the training set is often less than the dimensionality of images. In such cases, the traditional LDA algorithm fails as the within-class scatter matrix becomes singular. Secondly, computational difficulty occurs while working with such high-dimensional image-vectors. In high-dimensional and singular cases, LDA is often performed on a PCA transformed space. In this method each image is subjected to curvelet transform in order to extract the intrinsic features of the facial images. The selected curvelet subimages are then projected on eigenspace using PCA. Now on this PCA transformed space, LDA is performed; where there is no possibility of the within-scatter matrix to be degenerate. Then the resultant PCA+LDA features extracted from curvelet decomposed images form the final feature set for classification. The test images undergo a similar process and are converted to the same PCA+LDA representational basis. NN rule is used for identification.

### 5. Experimental results

#### 5.1. Datasets

##### 5.1.1. ORL database [41]

This database contains 10 different images of each of 40 distinct subjects taken between April 1992 and April 1994 at Olivetti Research Laboratory, Cambridge, UK. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The images are in '.pgm' format and of dimension  $92 \times 112$  (width  $\times$  height) 8-bit gray levels (Fig. 4).

##### 5.1.2. Essex Grimace database [42]

A sequence of 20 images each for 18 individuals consisting of male and female subjects was taken, using a fixed camera. During the sequence the subjects move their head and make grimaces which get more extreme



Fig. 4. Sample images from ORL.



Fig. 5. Sample images from Essex Grimace.



Fig. 6. Sample images from Yale face.

towards the end of the sequence. There is about 0.5 s between successive frames in the sequence. Images are taken against a plain background, with very little variation in illumination. The images are in '.jpg' format and of size  $180 \times 200$  (Fig. 5).

### 5.1.3. Yale face database [43]

The Yale face database contains 165 grayscale images in GIF format ( $320 \times 243$ ) of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/ no glasses, normal, right-light, sad, sleepy, surprised, and wink (Fig. 6).

## 5.2. Implementation and experimental results

To design the experiments, we randomly selected five images per subject for ORL database, eight images per subject for Essex Grimace and six images per subject for Yale face database as prototype and used the rest for testing. A five-fold cross-validation has been performed in each case. The color images of Essex Grimace dataset have been converted to gray-scale images, as color information does not have much contribution in this edge-based method. As a preprocessing step, the images have been reduced to half of their size. ORL images have not been resized. For curvelet based PCA and PCA+LDA framework, resized images are decomposed using curvelet transform at scale = 3 and angle = 8. For extraction of curvelet discriminant features, scale value is changed to 4, in order to avoid within-scatter matrix singular cases as mentioned in Section 4.3. This produces 33 components, including 1 approximate and 32 detailed subbands. We have worked with two subbands only. To further reduce

the dimensionality of feature vectors PCA/LDA/PCA+LDA is applied on these approximate components. After the curvelet sub-images are projected to desired feature-space, a simple distance based (L1 Norm) NN classifier is employed to perform the identification task. Number of principal components is varied to display how recognition rate changes with selection of eigenvectors. The recognition rates for these three approaches are shown in Figs. 7, 9 and 11. These results have been achieved by averaging the recognition rates of five different rounds of face recognition. However, the imaging system is not perfect and facial images can be noisy. Hence evaluation of our system under various noise levels. Noise with mean 0 and variance 0–0.2 has been added to the input images and recognition error rate has been graphed against noise level for three subspaces, considering 60 components each time. Results under noisy condition are shown in Figs. 8, 10 and 12.

### 5.3. Comparative study

In Section 5.2, different results of curvelet based PCA/LDA/PCA+LDA techniques have been presented. In order to establish the reliability of the proposed methods we have compared them against well-established existing techniques like standard eigenface based methods and wavelet based methods. For wavelet based schemes, a 3-level wavelet decomposition using standard 'Haar' wavelet (as it is often used in wavelet based face recognition methods) was performed. Feature size is 60 for each case; ratio of training and test images remains same as before. We have duplicated some recently developed techniques to compare with curvelet based results: wavelet based kernel Associative Memory (kAM) [44], wavelet based weighted

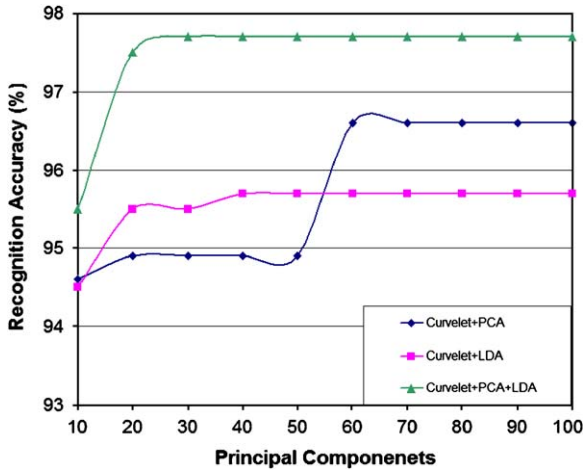


Fig. 7. Curvelet based results for ORL.

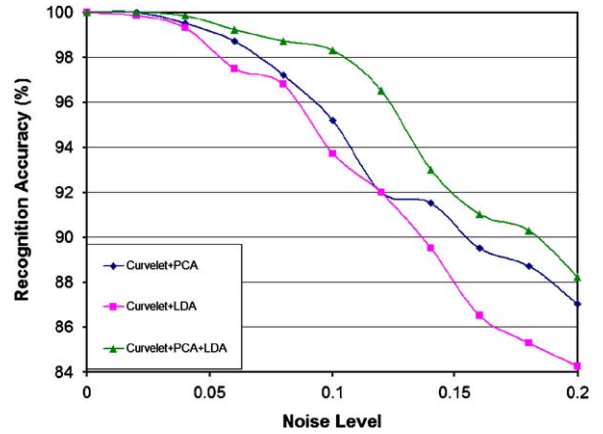


Fig. 10. System performance under noisy condition (Essex Grimace).

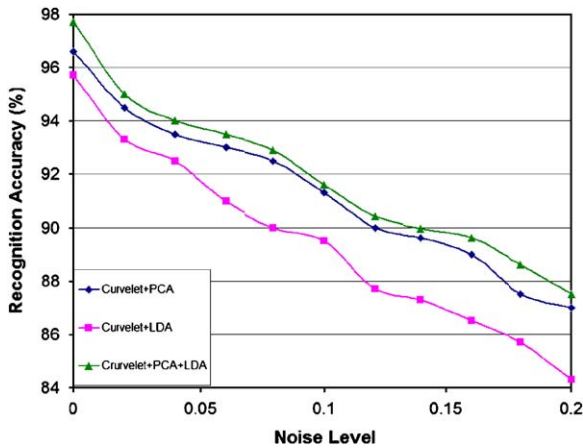


Fig. 8. System performance under noisy condition (ORL).

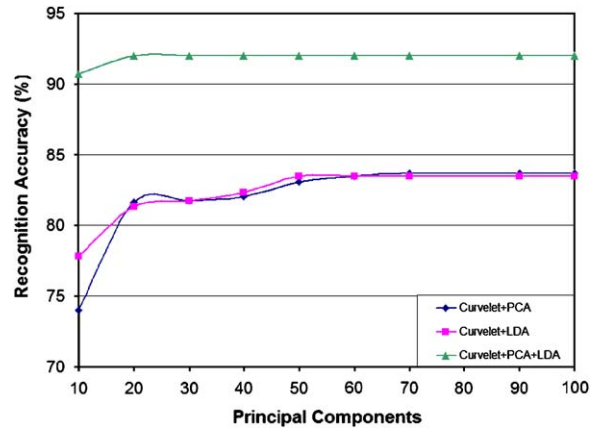


Fig. 11. Curvelet based results for Yale.

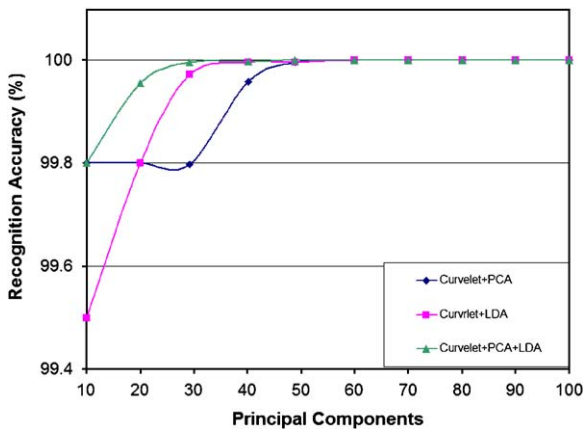


Fig. 9. Curvelet based results for Essex Grimace.

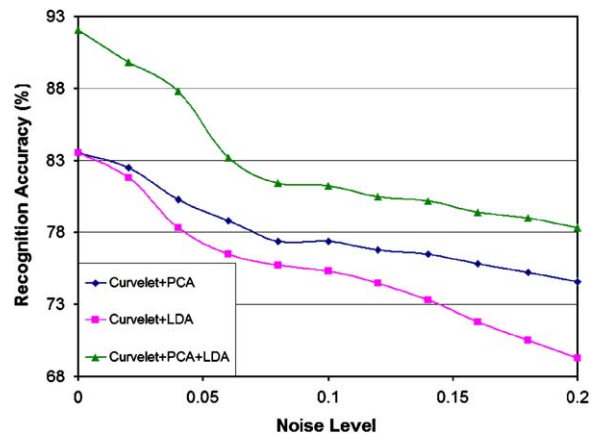


Fig. 12. System performance under noisy condition (Yale).

modular PCA [45] and discriminant waveletfaces using Nearest Feature Line (NFL) classifier [46]. Recognition rates reported in Table 1 are achieved after a five-fold cross-validation.

**Table 1**  
Comparative study.

Method	Recognition accuracy (%)		
	ORL	Essex	Yale
Standard eigenface [2]	92.2	69.4	76.0
Waveletface [3]	92.5	92.5	83.3
Curveletface	94.5	97.2	82.6
Waveletface+PCA [3]	94.5	98.5	84.0
Waveletface+LDA [46]	94.7	100	84.6
Waveletface+weighted modular PCA [45]	95.0	98.4	83.6
Waveletface+LDA+NFL [46]	95.2	100	83.5
Curveletface+LDA	95.6	100	83.5
Waveletface+kAM [44]	96.6	100	84.0
Curveletface+PCA	96.6	100	83.9
Curveletface+PCA+LDA	97.7	100	92.0

#### 5.4. Discussion

Various results presented in Sections 5.2 and 5.3 demonstrate that high recognition accuracy can be achieved using curvelet subspaces for face recognition. It seems that for Essex Grimace database (Fig. 9) recognition rate is almost perfect for all curvelet based methods, indicating the robustness of curvelet based approaches against extreme expression variation. Our techniques have been found to be robust against significant variation in illumination and facial details (present in ORL and Yale) as well. For all the databases, most prominently for Yale database (Fig. 11), PCA-LDA framework shows significant performance improvement over the PCA based and LDA based schemes. Though, LDA is usually expected to do better than PCA, it is not reflected in our experimental study as we have worked with small databases. LDA is particularly suited for high-dimensional data and can be outperformed by PCA otherwise. When compared to standard techniques like eigenface and waveletface based methods, curvelets show promise. Even simple curvelet features alone can outperform standard eigenface and waveletface. Curveletface and PCA-LDA framework works the best and shows 5.5% and 5.2% gain in accuracy over eigenface and waveletface for ORL. In [33] an average recognition accuracy of 92% for ORL database is reported. Using training and test sets of same size our curvelet-PCA framework shows an improvement of about 4.6% and PCA-LDA scheme shows 5.7% performance improvement over the proposed method in [33]. In case of Yale dataset, results of curvelet subspaces and wavelet subspaces are comparable, though curvelet based PCA-LDA scheme is able to achieve almost 8% higher matching rate. In Section 3 we argued that since curvelets are capable of capturing the crucial edge information, they can perform competently as the bases of face recognition. By comparing the results of curvelets with wavelets, its nearest cousin, we can definitely prove our point—curvelets are an efficient or even better alternative to wavelets. However, the trade off is, wavelet features are faster to compute, while curvelets can provide high accuracy. It has been found empirically, that in a standard personal computer with

2 GHz AMD processor and 1 GB RAM computation of wavelet features is almost 1.5 times faster than that of curvelet's using same image size and a decomposition level of 3. Increasing curvelet's scale or angle parameters can improve the accuracy, but has the overhead of extra computation time.

#### 6. Conclusion

The main contribution of this paper is the identification of three new feature spaces to address the problem of human face recognition from still images. These feature spaces are based on the PCA/LDA/PCA+LDA spaces of the features extracted by a new multiresolution analysis tool called digital curvelet transform. This is the first time curvelet subspaces are being explored. It is unlikely that the best possible results can be produced in this initial attempt. Rather, our work can be considered as a stepping stone for future works in this direction. Future work is suggested towards developing a complete curvelet based face recognition system and working on newly developed dimensionality reduction methods like 2DPCA, B-2DPCA. Performance of employing different classification methods can also be studied.

#### References

- [1] W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face recognition: a literature survey, *ACM Computing Surveys*, 2003, pp. 399–458.
- [2] M. Turk, A. Pentland, Face recognition using eigenfaces, in: *Proceedings of the Computer Vision and Pattern Recognition*, 1991, pp. 586–591.
- [3] G.C. Feng, P.C. Yuen, D.Q. Dai, Human face recognition using PCA on wavelet subband, *Journal of Electronic Imaging* 9 (2) (2000) 226–233.
- [4] B. Schoelkopf, A. J. Smola, K. R. Mueller, Kernel principal component analysis, in: *Lecture Notes in Computer Science*, 1997.
- [5] J. Yang, A.F. Frangi, J. Yang, D. Zhang, Z. Jin, KPCA Plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition, *IEEE Transactions on PAMI* 27 (2) (2005) 230–244.
- [6] M.S. Bartlett, H.M. Lades, T. Sejnowski, Independent component representation for face recognition, in: *Proceedings of the SPIE Symposium on Electronic Imaging: Science and Technology*, 1998, pp. 528–539.
- [7] P.N. Bellhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, *IEEE Transactions on PAMI* 19 (7) (1997) 711–720.
- [8] K.I. Kim, K. Jung, H.J. Kim, Face recognition using kernel principal component analysis, *IEEE Signal Processing Letters* 9 (2) (2002) 1558–2361.
- [9] Y. Pang, D. Tao, Y. Yuan, X. Li, Binary two-dimensional PCA, *IEEE Transactions on Systems, Man and Cybernetics* 38 (4) (2008) 1176–1180.
- [10] J. Yang, D. Zhang, A.F. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance based face representation and recognition, *IEEE Transactions on PAMI* 26 (1) (2004) 131–137.
- [11] Y. Pang, Y. Yuan, X. Li, Effective feature extraction in high dimensional space, *IEEE Transactions on Systems, Man and Cybernetics* 38 (6) (2008) 1652–1656.
- [12] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, H. Zhang, Human gait recognition with matrix representation, *IEEE Transactions on Circuits and Systems for Video Technology* 6 (7) (2006) 896–903.
- [13] D. Tao, X. Li, X. Wu, S.J. Maybank, General tensor discriminant analysis and Gabor features for gait recognition, *IEEE Transactions on PAMI* 29 (10) (2007) 1700–1715.
- [14] X. Li, S. Lin, S. Yuan, D. Xu, Discriminant locally linear embedding with high-order tensor data, *IEEE Transactions on Systems, Man and Cybernetics, Part B* 38 (2) (2008) 342–352.
- [15] M. Li, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, *Pattern Recognition Letters* 26 (5) (2005) 527–532.



- [16] K.C. Chung, S.C. Kee, S.R. Kim, Face recognition using principal component analysis of Gabor filter responses, in: Proceedings of the IEEE, 1999, pp. 53–57.
- [17] C.F. Chen, Y.S. Tseng, C.Y. Chen, Combination of PCA and wavelet transforms for face recognition on 2.5D images, in: Proceedings of the Image and Vision Computing, NZ, 2003, pp. 343–347.
- [18] G.Y. Tian, S. King, D. Taylor, S. Ward, Wavelet based normalization for face recognition, in: Proceedings of CGIM, 2003.
- [19] Y. Pang, Y. Yuan, X. Li, Gabor based region covariance matrices for face recognition, IEEE Transactions on Circuits and Systems for Video Technology 18 (7) (2008) 989–993.
- [20] M.N. Do, M. Vetterli, The finite ridgelet transform for image representation, IEEE Transactions on Image Processing 12 (1) (2003) 16–28.
- [21] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Transactions on Image Processing 14 (12) (2005) 2091–2106.
- [22] E.J. Candes, D.L. Donoho, Curvelets—A Surprisingly Effective Non-adaptive Representation for Objects with Edges, Vanderbilt University Press, Nashville, TN, 2000.
- [23] J.L. Starck, E.J. Candes, D.L. Donoho, The curvelet transform for image denoising, IEEE Transactions on Image Processing 11 (2000) 670–684.
- [24] E.J. Candes, D.L. Donoho, Curvelets, multiresolution representation and scaling laws, in: SPIE Wavelet Applications in Signal and Image Processing, 2001.
- [25] M. Manikandan, A. Saravanan, K.B. Bagan, Curvelet transform based embedded lossy image compression, in: Proceedings of ICSCN, 2007, pp. 274–276.
- [26] M. Choi, R.Y. Kim, M.G. Kim, The curvelet transform for image fusion, in: Proceedings of ISPRS Congress, Istanbul, 2004.
- [27] J.L. Starck, F. Murtagh, E.J. Candes, D.L. Donoho, Gray and color image contrast enhancement by the curvelet transform, IEEE Transactions on Image Processing 12 (2003) 706–717.
- [28] J. Starck, M.K. Nguyen, F. Murtagh, Deconvolution based on the curvelet transform, in: Proceedings of ICIP, 2003.
- [29] J.L. Starck, D.L. Donoho, E.J. Candes, Very high quality image restoration by combining wavelets and curvelets, in: Proceedings of SPIE, vol. 4478, 2003.
- [30] J.L. Starck, D.L. Donoho, E.J. Candes, Astronomical image representation by curvelet transform, Astronomy & Astrophysics 398 (2002) 785–800.
- [31] A. Majumdar, Bangla basic character recognition using digital curvelet transform, Journal of Pattern Recognition Research 2 (2007) 17–26.
- [32] T. Mandal, A. Majumdar, Q.M.J. Wu, Face recognition by curvelet based feature extraction, in: Proceedings of ICIAR, vol. 4633, 2007, pp. 806–817.
- [33] J. Zhang, Z. Zhang, W. Huang, Y. Lu, Y. Wang, Face recognition based on curvefaces, in: Proceedings of Natural Computation, 2007.
- [34] A. Majumdar, A. Bhattacharya, Face recognition by multi-resolution curvelet transform on bit quantized facial images, in: Proceedings of the Conference on Computational Intelligence and Multimedia Applications, vol. 2, 2007, pp. 209–213.
- [35] E.J. Candes, L. Demanet, D.L. Donoho, L. Ying, Fast discrete curvelet transform, in: SIAM Multiscale Modeling and Simulations, 2007.
- [36] E.J. Candes, What is curvelet, Notices of American Mathematical Society 50 (2003) 1402–1403.
- [37] J.L. Starck, Image Processing by the Curvelet Transform, PPT.
- [38] B.S. Manjunath, R. Chellappa, C.v.d. Malsburg, A feature based approach to face recognition, in: Proceedings of IEEE Computer Vision and Pattern Recognition, 1992, pp. 373–378.
- [39] K. Fukunaga, Statistical Pattern Recognition, Academic Press, New York, 1999.
- [40] J. Yang, J. Yang, Why can LDA be performed on PCA transformed space?, Pattern Recognition 36 (2003) 563–566.
- [41] <[http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att\\_faces.zip](http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att_faces.zip)>.
- [42] <<http://cswwww.essex.ac.uk/mv/allfaces/grimace.zip>>.
- [43] <<http://cswwww.essex.ac.uk/mv/allfaces/faces94.html>>.
- [44] B.-L. Zhang, H. Zhang, S. Sam Ge, Face recognition by applying wavelet subband representation and kernel associative memory, IEEE Transactions on Neural Networks 15 (1) (2004) 166–177.
- [45] M. Zhao, P. Li, Z. Liu, Face recognition based on wavelet transform weighted modular PCA, in: Proceedings of the Congress in Image and Signal Processing, 2008.
- [46] J.T. Chien, C.C. Wu, Discriminant waveletfaces and nearest feature classifiers for face recognition, IEEE Transactions on PAMI 24 (2) (2002) 1644–1649.