3

Curvelet Based Feature Extraction

Tanaya Guha[1] and Q. M. Jonathan Wu[2]
[1]*ECE, University of British Columbia, BC, Canada*
[2]*ECE, University of Windsor, ON, Canada*

1. Introduction

Designing a completely automatic and efficient face recognition system is a grand challenge for biometrics, computer vision and pattern recognition researchers. Generally, such a recognition system is able to perform three subtasks: face detection, feature extraction and classification. We'll put our focus on feature extraction, the crucial step prior to classification. The key issue here is to construct a representative feature set that can enhance system-performance both in terms of accuracy and speed.

At the core of machine recognition of human faces is the extraction of proper features. Direct use of pixel values as features is not possible due to huge dimensionality of the faces. Traditionally, Principal Component Analysis (PCA) is employed to obtain a lower dimensional representation of the data in the standard eigenface based methods [Turk and Pentland 1991]. Though this approach is useful, it suffers from high computational load and fails to well-reflect the correlation of facial features. The modern trend is to perform multiresolution analysis of images. This way, several problems like, deformation of images due to in-plane rotation, illumination variation and expression changes can be handled with less difficulty.

Multiresolution ideas have been widely used in the field of face recognition. The most popular multiresolution analysis tool is the Wavelet Transform. In wavelet analysis an image is usually decomposed at different scales and orientations using a wavelet basis vector. Thereafter, the component corresponding to maximum variance is subjected to 'further operation'. Often this 'further operation' includes some dimension reduction before feeding the coefficients to classifiers like Support Vector Machine (SVM), Neural Network (NN) and Nearest Neighbor. This way, a compact representation of the facial images can be achieved and the effect of variable facial appearances on the classification systems can also be reduced. The wide-spread popularity of wavelets has stirred researchers' interest in multiresolution and harmonic analysis. Following the success of wavelets, a series of multiresolution, multidimensional tools, namely contourlet, curvelet, ridgelet have been developed in the past few years. In this chapter, we'll concentrate on Digital Curvelet Transform. First, the theory of curvelet transform will be discussed in brief. Then we'll talk about the potential of curvelets as a feature descriptor, looking particularly into the problem of image-based face recognition. Some experimental results from recent scientific works will be provided for ready reference.

2. Curvelet Transform

Before getting started with curvelet transform, the reader is suggested to go through the theory of multiresolution analysis, especially wavelet transform. Once the basic idea of wavelets and multiresolution analysis is understood, curvelets will be easier to comprehend.

2.1 Theory and Implementation

Motivated by the need of image analysis, Candes and Donoho developed curvelet transform in 2000 [Candes and Donoho 2000]. Curvelet transform has a highly redundant dictionary which can provide sparse representation of signals that have edges along regular curve. Initial construction of curvelet was redesigned later and was re-introduced as Fast Digital Curvelet Transform (FDCT) [Candes et al. 2006]. This second generation curvelet transform is meant to be simpler to understand and use. It is also faster and less redundant compared to its first generation version. Curvelet transform is defined in both continuous and digital domain and for higher dimensions. Since image-based feature extraction requires only 2D FDCT, we'll restrict our discussion to the same.

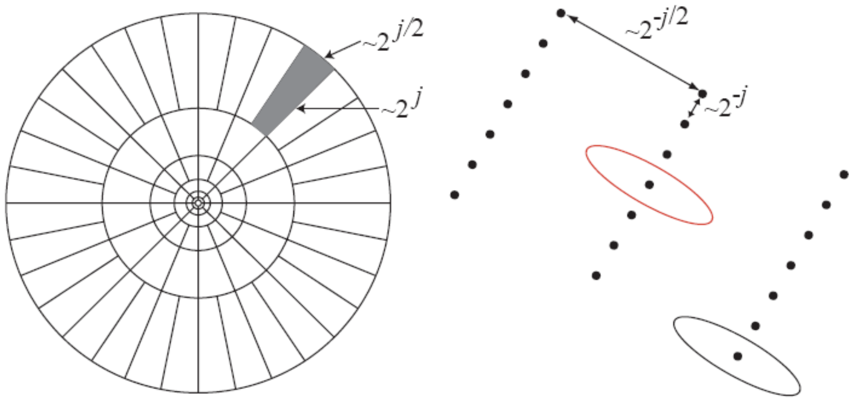
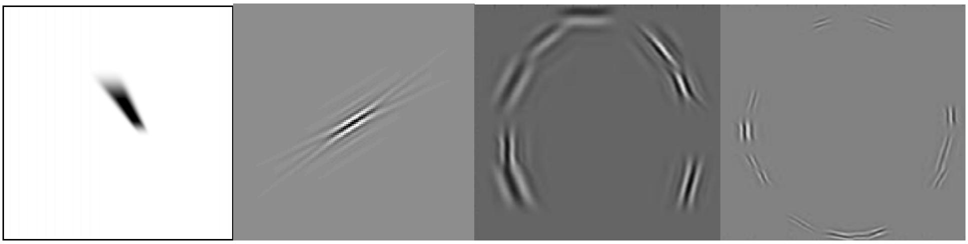


Fig. 1. Curvelets in Fourier frequency (left) and spatial domain (right) [Candes et al. 2006].

In order to implement curvelet transform, first 2D Fast Fourier Transform (FFT) of the image is taken. Then the 2D Fourier frequency plane is divided into wedges (like the shaded region in fig. 1). The parabolic shape of wedges is the result of partitioning the Fourier plane into radial (concentric circles) and angular divisions. The concentric circles are responsible for the decomposition of an image into multiple scales (used for bandpassing the image at different scale) and the angular divisions partition the bandpassed image into different angles or orientations. Thus if we want to deal with a particular wedge we'll need to define its scale j and angle ℓ. Now let's have a look at the spatial domain (fig. 1 right). Each of the wedges here corresponds to a particular curvelet (shown as ellipses) at a given scale and angle. This indicates that the inverse FFT of a particular wedge if taken, will determine the curvelet coefficients for that scale and angle. This is the main idea behind the implementation of curvelet transform. Figure 1 (right) represents curvelets in spatial Cartesian grid associated with a given scale and angle.



(a) (b) (c) (d)

Fig. 2. (a) a real wedge in frequency domain, (b) corresponding curvelet in spatial domain [Candes et al. 2006], (c) curvelets aligned along a curve at a particular scale, (d) curvelets at a finer scale [Starck et al. 2002].

There are two different digital implementations of FDCT: Curvelets via USFFT (Unequally Spaced Fast Fourier Transform) and Curvelets via Wrapping. Both the variants are linear and take as input a Cartesian array to provide an output of discrete coefficients. Two implementations only differ in the choice of spatial grid to translate curvelets at each scale and angle. FDCT wrapping is the fastest curvelet transform currently available [Candes et al. 2006].

Though curvelets are shown to form the shape of an ellipse in fig. 1, looking at fig. 2 (b-d), we can understand that actually it looks more like elongated needles. This follows from the parabolic scaling law (length $\approx$ width^2) that curvelets obey. The values of curvelet coefficients are determined by how they are aligned in the real image. The more accurately a curvelet is aligned with a given curve in an image, the higher is its coefficient value. A very clear explanation is provided in figure 3. The curvelet named '**c**' in the figure is almost perfectly aligned with the curved edge and therefore has a high coefficient value. Curvelets '**a**' and '**b**' will have coefficients close to zero as they are quite far from alignment. It is well-known that a signal localized in frequency domain is spread out in the spatial domain or vice-versa. A notable point regarding curvelets is that, they are better localized in both frequency and spatial domain compared to other transforms. This is because the wedge boundary is smoothly tapered to avoid abrupt discontinuity.

2.2 Comparison with wavelets

Fourier series requires a large number of terms to reconstruct a discontinuity within good accuracy. This is the well-known Gibbs phenomenon. Wavelets have the ability to solve this problem of Fourier series, as they are localized and multiscale. However, though wavelets do work well in one-dimension, they fail to represent higher dimensional singularities (especially curved singularities, wavelets can handle point singularities quite well) effectively due to limited orientation selectivity and isotropic scaling. Standard orthogonal wavelet transform has wavelets with primarily vertical, horizontal and diagonal orientations independent of scale.

Curvelet transform has drawn much attention lately because it can efficiently handle several important problems, where traditional multiscale transforms like wavelet fait to act. Firstly,

Curvelets can provide a sparse representation of the objects that exhibit '*curve punctuated smoothness*' [Candes, 2003], i.e. objects those are smooth except along a general curve with bounded curvature. Curvelets can model such curved discontinuities so well that the representation becomes as sparse as if the object were not singular. From figure 4, we can have an idea about the sparsity and efficiency of curvelet representation of curved singularities compared to wavelets. At any scale j, curvelets provide a sparse representation $O(2^{j/2})$ of the images compared to wavelets' $O(2^j)$. If an image function f is approximated by largest m coefficients as $\hat{f}_m$, then the approximation errors are given by:
Fourier transform

$$\left\| f - \hat{f}_m^F \right\|^2 \propto m^{-1/2},\ m \to +\infty$$

Wavelet transform

$$\left\| f - \hat{f}_m^W \right\|^2 \propto m^{-1},\ m \to +\infty$$

Curvelet transform

$$\left\| f - \hat{f}_m^C \right\|^2 \propto m^{-2} \log(m^3),\ m \to +\infty$$

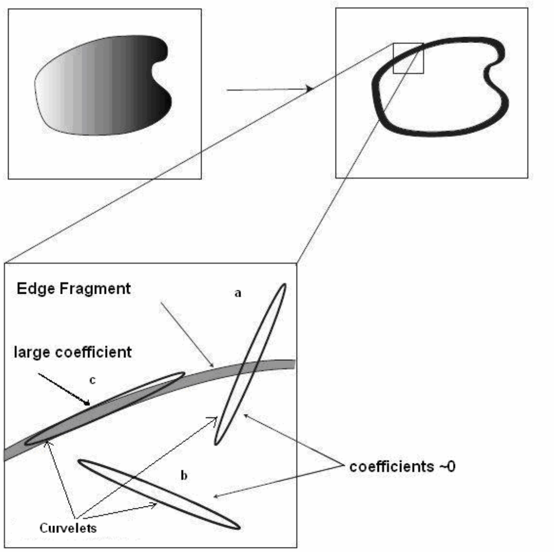


Fig. 3. Alignment of curvelets along curved edges [R]

The main idea here is that the edge discontinuity is better approximated by curvelets than wavelets. Curvelets can provide solutions for the limitations (curved singularity representation, limited orientation and absence of anisotropic element) the wavelet transform suffers from. It can be considered as a higher dimensional generalization of wavelets which have the unique mathematical property to represent curved singularities effectively in a non-adaptive manner.

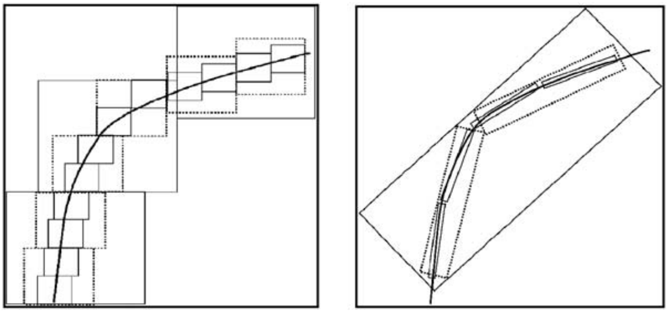


Fig. 4. Representation of curved sigularities using wavelets (left) and curvelets (right) [Starck, 2003].

2.3 Applications

Curvelet transform is gaining popularity in different research areas, like signal processing, image analysis, seismic imaging since the development of FDCT in 2006. It has been successfully applied in image denoising [Starck et al. 2002], image compression, image fusion [Choi et al., 2004], contrast enhancement [Starck et al., 2003], image deconvolution [Starck et al., 2003], high quality image restoration [Starck et al., 2003], astronomical image representation [Starck et al., 2002] etc. Examples of two applications, contrast enhancement and denoising are presented in figures 5 and 6. Readers are suggested to go through the referred works for further information on various applications of the curvelet transform. Recently, curvelets have also been employed to address several pattern recognition problems, such as face recognition [Mandal et al., 2007; Zhang et al., 2007] (discussed in detail in section 3), optical character recognition [Majumdar, 2007], finger-vein pattern recognition [Zhang et al., 2006] and palmprint recognition [Dong et al. 2005].

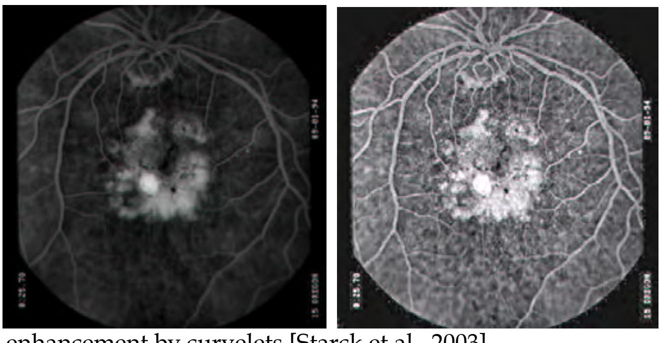


Fig. 5. Contrast enhancement by curvelets [Starck et al., 2003].

Fig. 6. Image denoising by curvelet [Starck et al. 2002].

3. Curvelet Based Feature Extraction for Faces

In the previous section, we have presented a theoritical overview of curvelet transform and explained why it can be expected to work better than the traditional wavelet transform. Facial images are generally 8 bit i.e. they have 256 graylevels. In such images two very close regions that have differing pixel values will give rise to edges; and these edges are typically curved for faces. As curvelets are good at approximating curved singularities, they are fit for extracting crucial edge-based features from facial images more efficiently than that compared to wavelet transform. We will now describe different face recognition methodologies that employ curvelet transform for feature extraction.

Typically, a face recognition system is divided into two stages: a training stage and a classification stage. In the training stage, a set of known faces (labeled data) are used to create a representative feature-set or template. In the classification stage, a unknown facial image is matched against the previously seen faces by comparing the features. Curvelet based feature extraction takes the raw or the preprocessed facial images as input. The images are then decomposed into curvelet subbands in different scales and orientations. Figure 7 shows the decomposition of a face image of size 112×92 (taken from ORL database) by curvelets at scale 2 (coarse and fine) and angle 8. This produces one approximate (75×61) and eight detailed coefficients (four of those are of size 66×123 and rest are of size 149×54). These curvelet decomposed images are called 'Curveletfaces'. The approximate curveletface contains the low-frequency components and the rest captures the high-frequency details along different orientations. It is sufficient to decompose faces using curvelet transform at scale 3 and angle 8 or 16. Increasing scales and/or orientations does not necessarily lead to significant improvement in recognition accuracy. If required, images can be reduced in size before subjecting them to feature extraction.

3.1 Curvelets and SVM

The first works on curvelet-based face recognition are [Zhang et al., 2007; Mandal et al. 2007]. A simple application of curvelet transform in facial feature extraction can be found in [Zhang et al., 2007]. The authors have used SVM classifier directly on the curvelet decomposed faces. The curvelet based results have been compared with that of wavelets.

Mandal et al. have performed 'bit quantization' before extracting curvelet features. The original 8 bit images are quantized to their 4 bit and 2 bit versions, as shown in figure 8. This is based on the belief that on bit quantizing an image, only bolder curves will remain in the lower bit representations, and curvelet transform will be able to make the most out of this curved edge information. During training, all the original 8 bit gallery images and their two bit-quantized versions are decomposed into curvelet subbands. Selected curvelet coefficients are then separately fed to three different Support Vector Machine (SVM) classifiers. Final decision is achieved by fusing results of all SVMs. The selection of the curvelet coefficients is done on the basis of their variance. The recognition results for these two methods are shown below.

Average Recognition Accuracy	Curvelet + SVM	Wavelet + SVM
	90.44 %	82.57%

Table 1. Face recognition results for ORL database [Zhang et al., 2007]

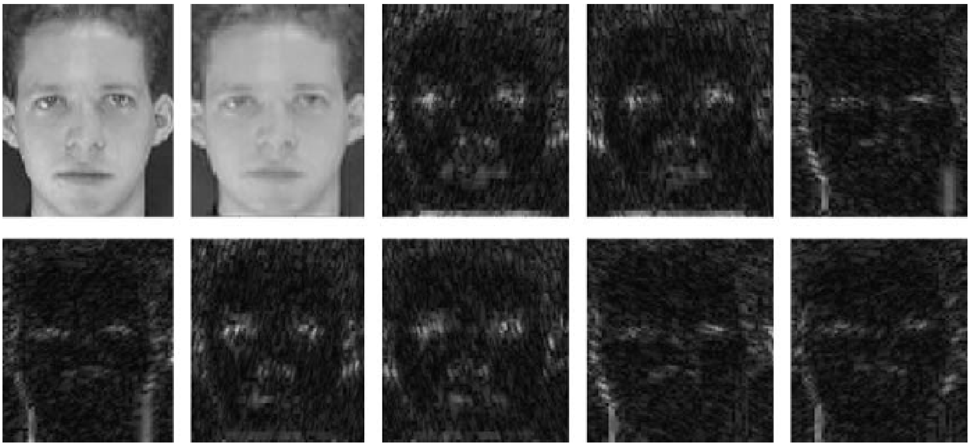


Fig. 7. Curvelet decomposition of a facial image - 1st image in the first row is the original image, 2nd image in the first row is the approximate coefficients and others are detailed coefficients at eight angles (all the images are resized to same dimension for the purpose of illustration only) [Mandal et al., 2009].

Fig. 8. Bit quantization: left most is the original 8 bit image (from ORL database), next two are 4 bit and 2 bit representations respectively [Mandal et al., 2007].

No. of Bits in Image	Accuracy of each Classifier	Accuracy after majority Voting	Rejection Rate	Incorrect Classification rate
8	96.9	98.8	1.2	0
4	95.6			
2	93.7			

Table 2. Recognition result for bit-quantized, curvelet decomposed images for ORL database [Mandal et al., 2007].

3.1 Curvelets and dimensionality reduction

However, even an image of size 64×64 when decomposed using curvelet transform at scale 3 (coarse, fine, finest) and angle 8 will produce the coarse subband of size 21×21 and 24 detailed coefficients of slightly larger size. Working with such large number of features is extremely expensive. Hence it is important to find a representative feature set. Only important curvelet subbands are selected depending on the amount of total variance they account for. Then dimensionality reduction methods like PCA, LDA and a combined PCA-LDA framework have been applied on those selected subbands to get an even lower dimensional representation [Mandal et al., 2009]. This not only reduces computational load, but also increases recognition accuracy.

The theory of PCA/LDA and will not be discussed here. Readers are requested to consult any standard book and the classical papers of Cootes et al. and Belhumeur et al. to understand the application of PCA and LDA in face recognition. PCA has been successfully applied on wavelet domain for face recognition by Feng et al. PCA has been employed on curvelet decomposed gallery images to form a representational basis. In the classification phase, the query images are subjected to similar treatment and transformed to the same representational basis. However, researchers argue that PCA, though is able to provide an efficient lower dimensional representation of the data, suffers from higher dimensional load and poor discriminative power. This issue can be resolved by the application of LDA that can maximize the within-class dissimilarity, simultaneously increasing the between-class similarity. This efficient dimensionality reduction tool is also applied on curvelet coefficients to achieve even higher accuracy and lower computational load. Often, the size of the training set is less than the dimensionality of the images. In such cases LDA fails to work, since the within-class scatter matrix become singular. Computational difficulty also arises while working with high-dimensional image vectors. In such high-dimensional and singular cases PCA is performed prior to LDA. Curvelet subimages are projected onto PCA-space and then LDA is performed on this PCA-transformed space. Curvelet features thus extracted are also robust against noise. These curvelet-based methods are compared to several existing techniques in terms of recognition accuracy in table 3. Though LDA is expected to work better than PCA that is not reflected in figures 9 and 10. This is because ORL is a small database and PCA can outperform LDA in such cases. In a recent work [Mohammed et al., 2009] Kernal PCA has been used for dimensionality reduction of curvelet features and even higher accuracy is achieved.

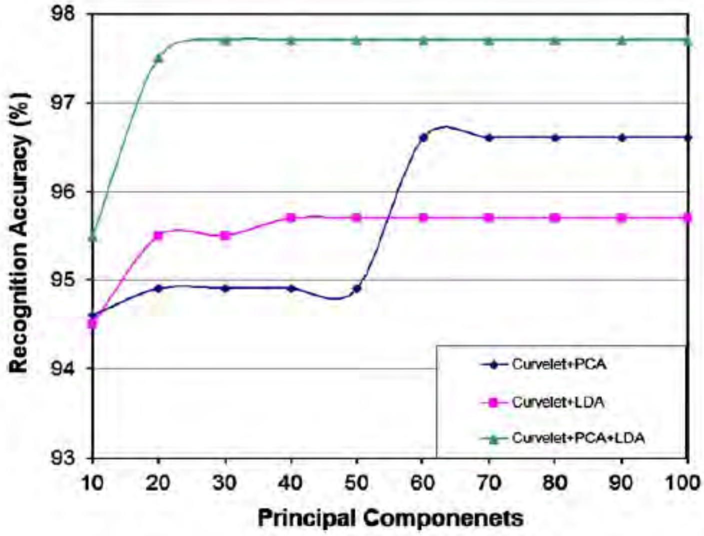


Fig. 9. Curvelet –based recognition accuracy for ORL database [Mandal et al., 2009]

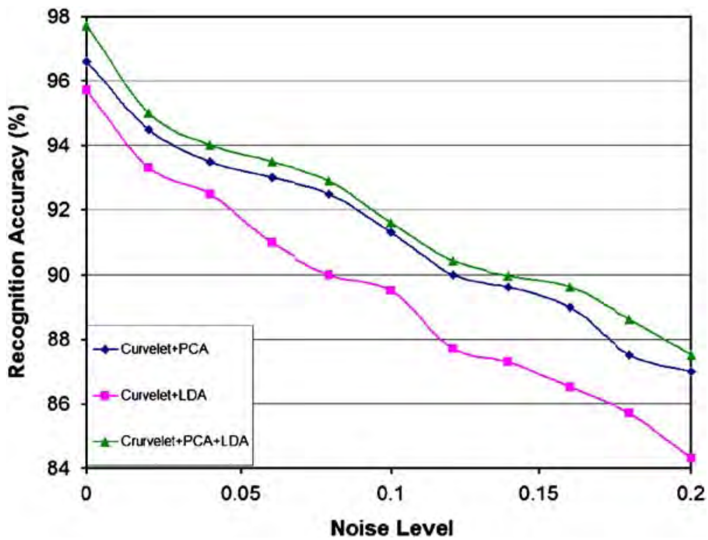


Fig. 10. Performance of curvelet-based methods against noise [Mandal et al, 2009]

Method	Recognition Accuracy (%)
Standard eigenface [Turk et al., 1991]	92.2
Waveletface [Feng et al.]	92.5
Curveletface	94.5
Waveletface + PCA [Feng et al., 2000]	94.5
Waveletface + LDA [Chien and Wu, 2002]	94.7
Waveletface + weighted modular PCA [Zhao et al., 2008]	95.0
Waveletface + LDA + NFL [Chien and Wu, 2002]	95.2
Curveletface + LDA	95.6
Waveletface + kAM [Zhang et al. 2004]	96.6
Curveletface + PCA	96.6
Curveletface + PCA + LDA	97.7

Table 3. Comparative study [Mandal et al., 2009]

4. Conclusion

In this chapter, newly developed curvelet transform has been presented as a new tool for feature extraction from facial images. Various algorithms are discussed along with relevant experimental results as reported in some recent works on face recognition. Looking at the results presented in tables 1, 2 and 3, we can infer that curvelet is not only a successful feature descriptor, but is superior to many existing wavelet-based techniques. Results for only one standard database (ORL) are listed here; nevertheless, work has been done on other standard databases like, FERET, YALE, Essex Grimace, Georgia-Tech and Japanese facial expression datasets. From the results presented in all these datasets prove the superiority of curvelets over wavelets for the application of face recognition. Curvelet features thus extracted from faces are also found to be robust against noise, significant amount of illumination variation, facial details variation and extreme expression changes.

The works on face recognition using curvelet transform that exist in literature are not yet complete and do not fully understand the capability of curvelet transform for face recognition; hence, there is much scope of improvement in terms of both recognition accuracy and curvelet-based methodology.

5. References

Candes, E. J. (2003). What is curvelet, *Notices of American Mathematical Society* vol. 50, pp. 1402–1403, 2003.

Candes, E. J.; Demanet, L.; Donoho, D. L. & Ying, L. (2007). Fast discrete curvelet transform, *SIAM Multiscale Modeling and Simulations*, 2007.

Candes, E. J. & Donoho, D. L. (2000). *Curvelets – A surprisingly effective non- adaptive representation for objects with Edges*, Vanderbilt University Press, Nashville, TN, 2000.

Chien, J. T. & Wu, C. C. (2002). Discriminant waveletfaces and nearest feature classifiers for face recognition, *IEEE Transactions on PAMI*, vol. 24, pp. 1644–1649, 2002.

Choi, M.; Kim, R. Y. & Kim, M. G. (2004). The curvelet transform for image fusion, *Proc ISPRS Congress*, Istanbul, 2004.

Dong, K.; Feng, G. & Hu, D. (2005). Digital curvelet transform for palmprint recognition, *Lecture notes in Computer Science, Springer*, vol. 3338, pp. 639-645, 2005.

Feng, G. C.; Yuen, P. C. & Dai, D. Q. (2000). Human face recognition using PCA on wavelet subband, *Journal of Electronic Imaging*, vol. 9(2), pp. 226–233, 2000.

Majumdar, A. (2007). Bangla basic character recognition using digital curvelet transform, *Journal of Pattern Recognition Research*, vol. 2, pp. 17-26, 2007.

Mandal, T.; Majumdar, A. & Wu, Q. M. J. (2007). Face recognition by curvelet based feature extraction. *Proc. International Conference on Image Analysis and Recognition*, vol. 4633, 2007, pp. 806–817 , Montreal, August, 2007.

Mandal, T.; Yuan, Y.; Wu Q. M. J. (2009). Curvelet based face recognition via dimensional reduction. *Signal Processing*, vol. 89, issue 12, pp. 2345-2353, December, 2009.

Mohammed, A. A.; Minhas, R.; Wu, Q. M. J. & Sid-Ahmed, M. A. (2009). A novel technique for human face recognition using non-linear curvelet feature subspace, *Proc. International Conference on Image Analysis and Recognition*, vol. 5627, July, 2009.

Starck, J. L. (2003). Image Processing by the Curvelet Transform, PPT.

Starck, J. L.; Candes, E. J. & Donoho, D. L. (2000). The curvelet transform for image denosing, *IEEE Transactions on Image Processing*, vol. 11, pp. 670–684, 2000.

Starck, J. L.; Donoho, D. L. & Candes, E. J. (2003). Very high quality image restoration by combining wavelets and curvelets, *Proceedings of SPIE*, vol. 4478, 2003.

Starck, J. L.; Donoho, D. L. & Candes, E. J. (2002). Astronomical image representation by curvelet transform, *Astronomy & Astrophysics*, vol. 398, pp. 785–800, 2002.

Starck, J. L.; Murtagh, F.; Candes, E. J. & Donoho, D. L. (2003). Gray and color image contrast enhancement by the curvelet transform, *IEEE Transactions on Image Processing*, vol. 12, pp. 706–717, 2003.

Starck, J. L.; Nguyen, M. K. & Murtagh, F. (2003). Deconvolution based on the curvelet transform, *Proc. International Conference Image Processing*, 2003.

Turk, M. & Pentland, A. Face recognition using eigenfaces (1991). *Proc. Computer Vision and Pattern Recognition*, pp. 586–591, 1991.

Zhang, J.; Ma, S. & Han, X. (2006). Multiscale feature extraction of finger-vein patterns based on curvelets and local interconnection structure neural network, *Proc. ICPR*, vol. 4, pp. 145-148, 2006.

Zhang, J.; Zhang, Z.; Huang, W.; Lu, Y. & Wang, Y. (2007). Face recognition based on curvefaces, *Proc. Natural Computation*, 2007.

Zhang, B. L.; Zhang, H. & Ge, S. S. (2004). Face recognition by applying wavelet subband representation and kernel associative memory, *IEEE Transactions on Neural Networks*, vol. 15 (1), pp. 166–177, 2004.

Zhao, M.; Li, P. & Liu, Z. (2008). Face recognition based on wavelet transform weighted modular PCA, *Proc. Congress in Image and Signal Processing*, 2008.