A Novel Technique for Human Face Recognition Using Nonlinear Curvelet Feature Subspace

Abdul A. Mohammed, Rashid Minhas, Q.M. Jonathan Wu, and Maher A. Sid-Ahmed

Department of Electrical Engineering,
University of Windsor, Ontario, Canada
{mohammea,minhasr,jwu,ahmed}@uwindsor.ca

Abstract. This paper proposes a novel human face recognition system using curvelet transform and Kernel based principal component analysis. 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 principal component analysis for obtaining an enhanced representative feature set. In this work we propose to use an improved scheme using kernel based principal component analysis (KPCA) for a comprehensive feature set generation. KPCA performs a nonlinear principal component analysis (PCA) using an integral kernel operator function and obtains features that are more meaningful than the ones extracted using a linear PCA. Extensive experiments were performed on a comprehensive database of face images and superior performance of KPCA based human face recognition in comparison with state-of-the-art recognition is established.

1 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.

Developing a face recognition model is quite 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 [2-3]. In the constituent based approach, recognition is achieved based on the relationship between human facial features such as eyes, nose, mouth and facial boundary [4-5]. The success of this

M. Kamel and A. Campilho (Eds.): ICIAR 2009, LNCS 5627, pp. 512–521, 2009.

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.

Face based approaches [1,6-7] 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) [7-8] has been proven to be an effective face based approach. 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 accuracy with a considerable reduction in computation.

Wavelet based approach using PCA for human face recognition [9] proposed by Feng *et al* has utilized a midrange frequency subband for PCA representation and has achieved improved accuracy and class separability. In a recent work, Mandal *et al* [10] 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. In this paper we propose to use coarse level curvelet coefficients together with a kernel based principal component analysis (KPCA) for face recognition. Experimental results on five well known face databases demonstrate that dimensionally reduced curvelet coefficients using KPCA offer better recognition in comparison with PCA based curvelet coefficients.

The remainder of the paper is divided into 4 sections. Section 2 discusses the curvelet transform, its variants along with their implementation details followed by a discussion of kernel based PCA in section 3. The proposed methodology is described in section 4. Experimental results are discussed in section 5 followed by conclusion, acknowledgment and references.

2 Curvelet Transform Feature Extraction

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. 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; contourlets and ridgelets.

Curvelet transform [11] 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 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.

2.1 Continuous -Time Curvelet Transform

Since the introduction of curvelet transform researchers have developed numerous algorithmic strategies [12] 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, non-negative, 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 by [11]

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

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 will always obey the admissibility conditions:

$$\begin{aligned} &\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \quad r > 0 . \\ &\sum_{l=-\infty}^{\infty} V^2(t - l) = 1, \quad t \in \Re . \end{aligned} \tag{2}$$

We define curvelets (as function of $x=(x_1,x_2)$) at a scale 2^{-j}, orientation θ_l, and position $x_k^{(j,l)} = R_{\theta}^{-1}(k_1.2^{-j}, k_2.2^{-j/2})$ 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 $f \in L^2(R^2)$ and a curvelet $\varphi_{j,k,l}$,

$$c(j,k,l) = \langle f, \varphi_{j,k,l} \rangle = \int_{R^2} f(x)\ \overline{\varphi_{j,k,l}}\ dx . \tag{3}$$

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:

$$\varphi_{j_0,k}(x) = \varphi_{j_0}(x - 2^{-j_0} k), \quad \hat{\varphi}_{j_0}(\omega) = 2^{-j_0}\, W_0(2^{-j_0}|\omega|) . \tag{4}$$

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 $(\phi_{j_0,k})_k$. Fig. 1 summarizes the key components of the constructions. The figure on the left represents the induced tiling of the frequency plane. In Fourier space, curvelets are supported near a parabolic wedge. Shaded area in left portion of fig.1 represents a generic wedge. The figure 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 an integral over the frequency plane as:

$$c(j,k,l)=\frac{1}{(2\pi)^2}\int \hat{f}(\omega)\overline{\hat{\varphi}_{j,k,l}(\omega)}\, d\omega=\frac{1}{(2\pi)^2}\int \hat{f}(\omega)U_j\left(R_{\theta_l}\omega\right)e^{i\left\langle x_k^{(j,l)},\omega\right\rangle}\, d\omega\ . \tag{5}$$

2.2 Fast Discrete Curvelet Transform

Two new algorithms have been proposed in [11] to improve previous implementations. New implementations of FDCT are ideal for deployment in large-scale scientific applications due to its 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.

2.2.1 FDCT via Wrapping [11]

1. Apply 2D FFT and obtain Fourier samples $\hat{f}[n_1,n_2]$, $-n/2 \le n_1,\ n_2 < n/2$.
2. For each scale j and angle ℓ, form the product $\tilde{U}_{j,\ell}[n_1,n_2]\hat{f}[n_1,n_2]$.
3. 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 n_1 and n_2 is now $0 \le n_1 < L_{1,j}$ and $0 \le n_2 < L_{2,j}$.
4. Apply the inverse 2D FFT to each $\tilde{f}_{j,\ell}$, hence collecting the discrete coefficients.

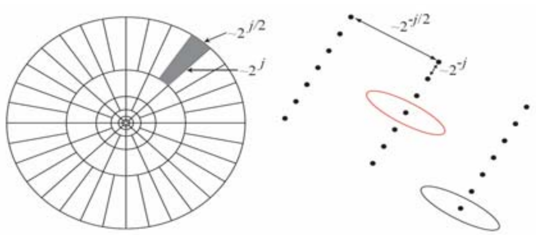

Fig. 1. Curvelet tiling of space and frequency [11]

In this work curvelet based features of human faces are extracted using FDCT via the wrapping technique. Coarse level coefficients are selected for face representation and their dimensionality is reduced using kernel based principal component analysis. Approximate coefficients are selected since they contain an overall structure of the image instead of high frequency detailed information which is insignificant and does not greatly impact the recognition accuracy. Fig. 2 shows an image from FERET database along with its approximate coefficients and detailed higher frequency coefficients at eight angles in the next coarsest level.

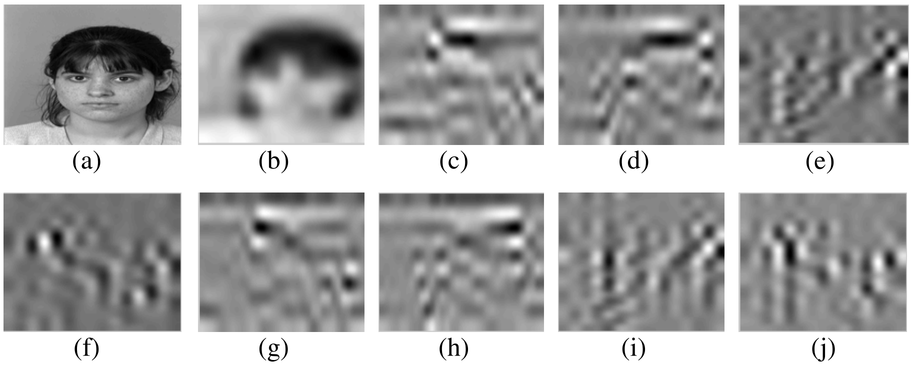

Fig. 2. (a) Original face image, (b) approximate curvelet coefficient, (c-j) 2^{nd} coarsest level curvelet coefficients at 8 varying angles

3 KPCA for Dimensionality Reduction

Principal component analysis (PCA) is a powerful technique for extracting structure information from higher dimension data. PCA is an orthogonal transformation of the coordinate system and is evaluated by diagonalizing the covariance matrix. Given a set of feature vectors $x_i \in R^N$, $i = 1,2,3......m$, which are centered with zero mean, their covariance matrix is evaluated as:

$$C = \frac{1}{m}\sum_{j=1}^{m} x_i x_j^T . \tag{6}$$

Eigenvalue equation, $\lambda v = Cv$ is solved where v is eigenvector matrix. To obtain a data with N dimensions, eigenvectors corresponding to the N largest eigenvalues are selected as basis vectors of the lower dimension subspace.

Kernel PCA is a generalization of PCA to compute the principal components of a feature space that is nonlinearly related to the input space. Feature space variables are obtained by higher order correlations between input variables. KPCA acts as a nonlinear feature extractor by mapping input space to a higher dimension feature space through a nonlinear map where the data is linearly separable. Cover's theorem [13] justifies the conversion of data to a higher dimensional space and formalizes the intuition that the number of separation increases with dimensionality and more views of the class and non class data are evident. Mapping achieved using the kernel trick solves the problem of nonlinear distribution of low level image features and acts as a dimensionality reduction step. Data is transformed from a lower dimension space to a higher dimension using the mapping function $\phi: R^N \to F$, and linear PCA is performed on F. The covariance matrix in the new domain is calculated using following equation:

$$\overline{C} = \frac{1}{m}\sum_{j=1}^{m} \phi(x_i)\ \phi(x_j)^T . \tag{7}$$

The problem is reduced to an Eigenvalue equation as in PCA and is solved using the identity $\lambda v = \overline{C}v$. As mentioned earlier the nonlinear map φ is not computed explicitly and is evaluated using the kernel function $K(x_i, x_j) = (\phi(x_i).\phi(x_j))$. The kernel function implicitly computes the dot product of vectors x_i and x_j in the higher dimension space. Kernels are considered as functions measuring similarity between instances. The kernel value is high if the two samples are similar and zero if they are distant. Some of the commonly used kernel functions and the mathematical equations associated with each kernel function are listed in Table 1.

Pairwise similarity amongst input samples is captured in a Gram matrix K and each entry of the matrix K_{ij} is calculated using the predefined kernel function $K(x_i,x_j)$. Eigenvalue equation in terms of Gram matrix is written as $m\lambda\beta = K\beta$.

K represents a positive semi definite symmetric matrix and contains a set of Eigenvectors which span the entire space. β denotes the column vector with entries $\beta_1, \beta_2........\beta_m$. Since the Eigenvalue equation is solved for β instead of eigenvector V_i of the kernel PCA, the entries of β are normalized in order to ensure that the eigen values of kernel PCA have unit norm in the feature space. After normalization the eigenvector matrix of kernel PCA is computed as $V=D\beta$ where $D=[\phi(x_1)\phi(x_2)......\phi(x_m)]$ is the data matrix in the feature space.

Table 1. Kernel Functions

Kernel Type	Mathematical Identity
Gaussian kernel	$k(xi, xj) = \exp\frac{(-\Vert x_i - x_j \Vert^2)}{2\alpha^2}$
Polynomial kernel	$k(xi, xj) = (x_i \,.\, x_j + \alpha)^d$, d = 1,2,3....
Sigmoid kernel	$\tanh\,(k(xi, xj) + \alpha)$

4 Proposed Method

Our proposed method deals with classification of face images using k-NN based classification utilizing reduced dimension feature vectors obtained from curvelet space. Images from each dataset are converted into gray level image with 256 gray levels. Conversion from RGB to gray level format along with a two fold reduction in the image size was the only pre-processing performed on the images. In addition to the mentioned adaptations there were no further changes made in an image that 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 datasets are of the same dimension, i.e. RxC. 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 better localization of higher frequency components with minimized aliasing effects. Vectorization is the next step to convert our curvelet coefficients into UxV dimension vector, called as curvelet vector, whereas UxV << RxC.

Applying k-NN classifier on curvelet vectors could be computationally expensive due to higher dimensionality of data originating from large image databases. Outliers and irrelevant image points being included into classification task can also effect the performance of our algorithm hence KPCA is implemented to reduce the dimension of curvelet vectors. KPCA was proposed in the pioneering work of [14] which computes principal components in a higher dimension feature space which is non-linearly related to the input space. Hence KPCA can reliably extract non-linear principal components while maintaining global content of the input space.

A Polynomial function based KPCA is used in our proposed method for dimensionality reduction to construct KPCA feature vectors. KPCA feature vectors retain the global structure of input space that facilitates accurate classification with lower computational complexity, diminished outliers and irrelevant information. Next, k-NN algorithm is trained using labeled KPCA feature vectors computed in step 5 (Table 2). We selected the k-NN based classification scheme due to its attractive properties and better performance in image-to-class scenario compared with other parametric classification schemes as argued by Boiman *et. al.* [15]. Finally, test image set feature vectors are classified using k-NN scheme utilizing Euclidean distance metric to compute the dissimilarity level between input images. Table 2 consists of detailed steps that demonstrate our proposed techniques.

Table 2. Main steps of our proposed classification scheme

INPUT: Randomly divide image dataset 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. Vectorize coarse level feature sets into *UxV* dimension vector.
3. Compute the kernel matrix K_{TR} and K_{TE} where each entry of the matrix is computed using a polynomial kernel function as mentioned in table - 1.
4. Solve eigenvalue equations:

   $$u \Lambda A_{TR} = K_{TR} A_{TR}$$

   $$u \Lambda A_{TE} = K_{TE} A_{TE}$$

   where Λ, A_x are eigenvalue and eigenvector matrices respectively.
5. Obtain kernel PCA based feature vectors, f_p, by computing principal component projections of each image into non-linear subspace using A_x.
6. Train k-NN algorithm; KPCA based feature vectors obtained in previous step using training images will be used.
7. Classify test images' feature vectors using k-NN trained in step 6.

5 Experimental Results

We tested our proposed method comprehensively using 5 distinctive faces databases namely, FERET, AT&T, Georgia Tech, Faces94 and JAFFE data sets. Before divulging into the experimental details and the results achieved using the proposed method, we will briefly describe the datasets used for face recognition.

5.1 Datasets

The FERET was sponsored by the Department of Defense in order to develop automatic face recognition capabilities that could be employed to assist security, intelligence and law enforcement personnel in the performance of their duties. The final corpus consists of 14051 eight-bit grayscale images of human heads with views ranging from frontal to left and right profiles.

AT&T face database contains 10 different images for each of 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, the faces were captured at different 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.

5.2 Results and Discussion

As described earlier image datasets are converted from RGB to gray level and image size is reduced by a factor of 2 in our experiments. In the FERET database different number of images exists for different subjects so 45% of images from each subject were used as prototypes and the remaining 55% for testing. Five images of each subject from AT&T database are randomly selected as prototype and the remaining 5 are used for testing the recognition accuracy. Similarly 9 images of each subject of the Georgia Tech dataset, 8 images of each subject from the Faces94 dataset and 9 images of each subject from the JAFFE dataset are randomly selected for training. Both the testing and training sets of images are decomposed using curvelet transform at 5 scales and 8 different angles. Amongst the curvelet coefficients only approximate coefficients are selected as feature vectors since they closely represent and approximate the input image. The selected feature vectors are dimensionally reduced with

KPCA using a 3rd degree polynomial kernel. k-NN is performed on dimensionally reduced feature vectors using different neighborhood size for classification and the results obtained with a neighborhood size of 5 are reported. The above process was repeated thrice for all databases and average results are tabulated. The recognition accuracy for each of the aforementioned databases using our proposed method is listed in Tables 3-5. Varying number of principal components are used to emphasize the recognition accuracy achieved using PCA and KPCA prior to saturation. It is clearly evident that the proposed method outperforms the conventional human face recognition accuracy obtained using conventional PCA based techniques.

Table 3. Recognition rates for FERET and AT&T database

Number of Components	**Recognition accuracy (%) for FERET face database**		**Recognition accuracy (%) for AT&T face database**	
	Curvelet + PCA [15]	Proposed Method	Curvelet + PCA [15]	Proposed Method
5	47.8495	66.6667	45.5	81
10	56.4516	82.2581	69	91
15	56.9892	86.0215	72.5	93
20	54.3011	88.7097	71.5	95.5
25	52.1505	88.7097	74.5	96

Table 4. Recognition rates for Georgia Tech face database

Number of Components	**Recognition accuracy (%) Curvelet + PCA [15]**	**Recognition accuracy (%) Proposed Method**
5	76.6667	96
10	79.6667	97
15	80.3333	97
20	81.3333	97.3333
25	81.6667	97.3333

Table 5. Recognition rates for Faces94 and JAFFE database

Number of Components	**Recognition accuracy for Faces94 face database**		**Recognition accuracy for JAFFE face database**	
	Curvelet + PCA [15]	Proposed Method	Curvelet + PCA [15]	Proposed Method
5	96.8202	98.6842	85.3846	93.8462
6	97.3684	98.9583	90.7692	97.6923
7	97.9715	99.1228	93.0769	96.1538
8	98.3004	99.1228	94.6154	96.9231
9	98.1360	99.2325	93.8462	100

6 Conclusion

We proposed 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. Kernel based PCA is utilized for dimension reduction and extraction of nonlinear feature sets. k-NN based scheme is employed for ascertaining recognition and classification. Experiments are performed using five popular human face databases and significant improvement in recognition accuracy is achieved. The proposed method considerably outperforms conventional face recognition systems using standard PCA.

References

1. Turk, M.A., Pentland, A.P.: Face Recognition using Eigenfaces. In: Proc. Computer Vision and Pattern Recognition, pp. 586–591 (1991)
2. Chow, G., Li, X.: Towards a system for automatic facial feature detection. Pattern Recognition 26(12), 1739–1755 (1993)
3. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.J.: Face Recognition: A Literature Survey. In: ACM Computing Surveys, pp. 399–458 (2003)
4. Goudail, F., Lange, E., Iwamoto, T., Kyuma, K., Otsu, N.: Face recognition system using local autocorrelations and multiscale integration. IEEE Trans. Pattern Anal. Mach. Intell. 18(10), 1024–1028 (1996)
5. Valentin, D., Abdi, H., O'Toole, A.J., Cottrell, G.W.: Connectionist models of face processing: A Survey. Pattern Recognition 27, 1209–1230 (1994)
6. Swets, D.L., Weng, D.L.: Using discriminant eigenfeatures for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell. 18(8), 831–836 (1996)
7. Kirby, M., Sirovich, L.: Application of the Karhunen-Loeve procedure for the characterizati-on of human faces. IEEE Trans. Pattern Anal. Mach. Intell. 12, 103–108 (1990)
8. Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, Seattle, pp. 84–91 (1994)
9. Feng, G.C., Yuen, P.C., Dai, D.Q.: Human Face Recognition using PCA on Wavelet Subband. Journal of Electronic Imaging 9(2), 226–233 (2000)
10. Mandal, T., Wu, Q.M.J.: Face Recognition using Curvelet Based PCA. In: ICPR (2008)
11. Candès, E.J., Demanet, L., Donoho, D.L., Ying, L.: Fast discrete curvelet transforms. In: Multiscale Model. Simul., pp. 861–899 (2005)
12. Starck, J.L., Elad, M., Donoho, D.L.: Redundant multiscale transforms and their application for morphological component analysis. Advances in Imaging and Electron Physics 132, 287–342 (2004)
13. Cover, T.M.: Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition. IEEE Transaction on Electronic Computers 14(3), 326–334 (1965)
14. Scholkopf, B., Smola, A., Muller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 12(5), 1299–1319 (1998)
15. Boiman, O., Shechtman, E., Irani, M.: In Defense of Nearest-Neighbour Based Image Classification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2008)