A face portion based recognition system using multidimensional PCA

A. Adeel Mohammed, Rashid Minhas, Q. M. Jonathan Wu, M. A. Sid-Ahmed
Department of Electrical and Computer Engineering
University of Windsor, Windsor, ON, N9B 3P4
(mohammea, minhasr, jwu, ahmed)@uwindsor.ca

***Abstract*— In this paper a new human face recognition algorithm based on localized face portion of an image is proposed. Extracted pure facial image is decomposed using curvelet transform and its selected subband is utilized for classification. Subband exhibiting a maximum standard deviation is dimensionally reduced using an improved dimensionality reduction technique, i.e., bidirectional two-dimensional principal component analysis to generate distinctive feature sets. These feature sets are used for training and testing an extreme learning machine classifier. Notable contributions of the proposed work include significant improvements in classification rate, speed and negligible dependence on the number of prototypes.**

I. Introduction

Face recognition has attracted research community during the last few decades as they are the most common visual patterns in our environment. Face recognition is non-intrusive i.e. images can be captured, identified or verified even without the knowledge and physical interaction of the subject. Moreover, an expert is not required to analyze and interpret the results and data can be easily collected with simple image acquisition devices.

Geometric feature-based face recognition is achieved by evaluating attributes such as distances and angles between various facial features, i.e., eyes, nose, mouth and chin [1]. Geometric methods depend upon feature extraction and measurement process and research has shown that face recognition based on finding local image features and inferring identity using geometric relations is often ineffective [2]. Appearance-based methods have emerged as fitting alternatives that use low-dimensional representation of face images to perform recognition. Global methods treat human face as a two-dimensional intensity variation pattern and establish recognition through detection and matching of statistical properties [3].

Kirby *et al.* [4] represented human faces as a linear combination of weighted eigenvectors using principal component analysis (PCA). PCA based face recognition systems suffer from poor discriminatory power and high computational load, and therefore to eliminate the shortcomings of standard PCA based systems, Bartlett *et al.* [5] proposed the use of independent component analysis (ICA). In [6] authors utilized linear discriminant analysis (LDA) to maximize the ratio of between-class scatter matrix and the within-class scatter matrix for improved face recognition.

Multiresolution analysis based schemes have been proposed to improve performance, deal with high image dimensionality, variations in viewpoint, illumination and facial expression. Face images are transformed into a new domain and later PCA and/or other dimensionality reduction techniques are employed. Some of the well known wavelet based face recognition architectures include wavelet based PCA [7], wavelet based LDA [8], wavelet based kernel association memory [9]. Emergence of curvelets has prompted researchers to apply them to several areas of image processing. Recent works in literature that are based curvelet transform include curvelet based PCA [10], curvelet based LDA [11] and curvelet based PCA+LDA [11]. Limitations of existing algorithms include large sensitivity to viewpoint variations and number of prototypes as well as slow classification speed.

In this work we propose a classification scheme that localizes the face and eliminates the background information from the image in a manner that the majority of the cropped image consists of the facial pattern. As a result of head movement some of the background information is involuntarily included in evaluating feature vectors. In Section II we briefly introduce curvelet transform, followed by a discussion of our proposed bidirectional two-dimensional principal component analysis (B2DPCA) in Section

978-1-61284-857-0/11/$26.00

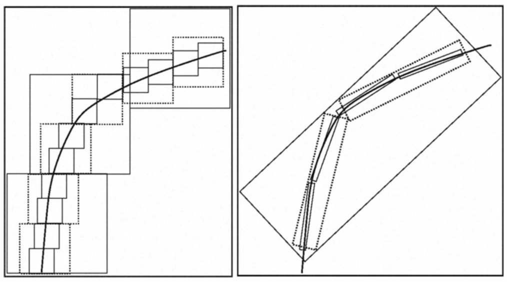

Fig. 1. Wavelets vs. Curvelets [12]

III. Section IV and V describe our proposed algorithm and experimental results, respectively.

II. CURVLET TRANSFORM

Curvelet transform [13] overcomes shortcomings of existing multiresolution analysis schemes and offers improved directional capacity to represent edges and other singularities along curves. Curvelets outperform wavelets in situations that require optimal sparse representation of objects with edges, representation of wave propagators, image reconstruction with missing data etc. Let us assume that a function f has a discontinuity across a curve, and is otherwise smooth as shown in Figure 1. Approximating f from the best m-terms in the Fourier expansion at a specified error rate in wavelet domain would require $O(m^{-1})$ terms, whereas a curvelet expansion demands only $O(m^{-2})$ terms. Curvelet transform is a multiscale non-standard pyramid with numerous directions and positions at each length and scale. Curvelets offer anisotropic and a locally adaptive scaling unlike other pyramid schemes.

Fast discrete curvelet transform (FDCT) requires an utmost 10 folds computational complexity as compared to fast fourier transform (FFT) operating on a similar sized data. We used FDCT via wrapping, proposed by authors in [14] for image analysis. Interested readers are requested to refer to [14] for additional mathematical details.

- Compute 2D FFT coefficients and obtain Fourier samples $\hat{f}[n_1, n_2]$ where $-n/2 < n_1$ and $n_2 < n/2$.
- For each scale j and angle l, form the product $\tilde{U}_{j,l}[n_1, n_2]\hat{f}[n_1, n_2]$
- Wrap this product around the origin and obtain $\tilde{f}_{j,l}[n_1, n_2] = W(\tilde{U}_{j,l}\hat{f})$ $[n_1, n_2]$, where the range of n_1, n_2 and θ respectively are $0 < n_1 < L_{1,j}$, $0 < n_1 < L_{2,j}$ and $(-\pi/4, \pi/4)$.
- Apply inverse 2D FFT to each $\tilde{f}_{j,l}$ and save discrete curvelet coefficients.

In the first two stages, Fourier frequency plane of the image is divided into radial and angular wedges owing to the parabolic relationship between a curvelets length and width. Each wedge corresponds to curvelet coefficient at a particular scale and angle. Step 3 is essentially required to re-index the data around the origin. Finally, inverse FFT is applied to collect discrete curvelet coefficients in the spatial domain.

III. BIDIRECTIONAL TWO-DIMENSIONAL PRINCIPAL COMPONENT ANALYSIS

Principal component analysis is a data representation technique widely used in pattern recognition and compression schemes. PCA fails to capture minor variance unless it is explicitly accounted in the training data. Wiskott *et.al* [15] proposed a bunch graph matching technique to overcome limitations and flaws of linear PCA. In [16], Yang *et al.* proposed two dimensional PCA for image representation. As opposed to PCA, 2DPCA is based on 2D matrices rather than 1D vectors. Therefore, image matrix does not need to be vectorized prior to feature extraction. Instead an image covariance matrix is directly computed using original image matrices.

Let X denote a q dimensional unitary column vector. To project a pxq image matrix A to X; linear transformation $Y = AX$ is used which results into a p dimensional projected vector Y. The total scatter is characterized by the trace of S_x, i.e., covariance matrix of the projected feature vectors; $J(X) = tr(S_x)$. $G_t = E[(A - EA)^T(A - EA)]$ is a nonnegative qxq image covariance matrix. If there are $\mathbb{M}$ training samples, the α^{th} image sample is denoted by $p \times q$ matrix A_α.

$$S_x = E[(A - EA)X][(A - EA)X]^T. \quad (1)$$

$$tr(S_x) = X^T[E(A - EA)^T(A - EA)]X. \quad (2)$$

$$J(X) = X^T G_t X. \quad (3)$$

Optimal projection axes X_{opt} represents a collection of M orthonormal eigen vectors $X_1, X_2, ..., X_M$ of G_t corresponding to M largest eigen values. Hence, dimensionality of every image A_α is reduced by post multiplying and pre-multiplying the image with optimal projection axes as $X_{opt}^T A_\alpha X_{opt}$.

A limitation of 2DPCA based recognition is its operability along row direction only. Our dimensionality reduction algorithm operates independently along row and column directions in order to better preserve the neighborhood relationship and to generate distinctive feature sets. Our proposed technique closely follows the work of [16] and generates an image covariance matrix $G_{t\alpha}$ and further optimizes it exploiting optimal project axes. Once optimal projection axes $X_{opt\alpha}$ is calculated, the dimensionality of every image A_α is reduced along its columns to generate new image sets A_β. The newly generated image sets are subsequently treated as a fresh database and a latest image covariance matrix $G_{t\beta}$ and optimal projection axes $X_{opt\beta}$ are evaluated. Finally, every new image A_β is pre-multiplied by $X^T_{opt\beta}$. Hence, unlike traditional 2DPCA, a two fold approach is adopted in our proposed B2DPCA algorithm to reduce image dimensionality.

$$A_\beta = A_\alpha X_{opt\alpha}. \quad A_\Theta = X^T_{opt\beta} A_\beta. \tag{4}$$

IV. Proposed Scheme

The proposed method is based on image decomposition of curvelet transform and uses dimensionally reduced coefficients for recognition. Distinctive feature sets generated using B2DPCA are used to train and test an extreme learning machine (ELM) classifier [17].

Images from each database are uniformly cropped in order to eliminate the background information and to ensure that the image predominantly contains facial pattern. In addition to facial features, some head information and background information is also retained due to head orientation. Cropped images are converted into gray level image with a two fold reduction in image size. Each database is randomly divided into training and testing set so that 40-45% of images of each subject are used as prototypes and remaining images are used during testing phase. Curvelet transform is used to generate initial feature vectors since it offers superior performance in presence of singularities in higher dimension, and enhances localization of higher frequency components with minimized aliasing effects. Input images are resized to $R \times C$, since analogous image sizes support generation of curvelet feature vectors with identical level of global information. Furthermore, curvelet decomposition of all images is computed at 3 scales and 8 angular orientations thus, generating 25 distinct subbands.

In contrast to the most recent work in literature [11] that uses two subbands, we have selected only one subband since the difference between standard deviations of the coarsest curvelet subband and the next coarser subband is quite significant. The proposed approach is based on selecting a subband with the utmost standard deviation which leads to momentous savings in computational cost during dimensionality reduction. The curvelet transform of all images in every dataset is evaluated and standard deviation of all the curvelet subbands is determined. It is noticed that approximate curvelet subband holds the maximum standard deviation amongst all 25 curvelet subbands.

B2DPCA is used to generate unique feature sets and to minimize computational complexity of our framework. Yang *et al.'s* [16] 2DPCA calculates a single covariance matrix to reduce the image dimensionality along its rows and columns respectively, whereas, in our proposed approach image dimensionality along orthogonal directions is reduced independently of each axis. The intermediate features are extracted by initially reducing dimension of initial feature matrix, i.e., selected curvelet subband along its columns. Later dimensionality of intermediate features is reduced along its rows in order to generate a final feature set, each of size $U' \times V'$, where $U' \times V' << U \times V$ (please refer to Section 3 for implementation details of B2DPCA). The modified approach used in this paper helps us to preserve critical neighborhood information between adjoining pixels and to generate distinctive features. For each dataset, dimensionally reduced curvelet feature sets are randomly selected for training of an ELM, whereas, remaining features of the same dataset are used to judge the separability of our framework. Please note that we do not assume any *a priori* knowledge of the face orientation and illumination conditions.

V. Results and Discussion

Experiments are performed using our proposed method using Sheffield [18] and ORL [19] face database. Sheffield face database consists of 564 images of 20 individuals. The database consist of images of individuals with mixed race, gender and appearance. Each individual is imaged in a range of poses from left/right profiles to frontal views with small angular rotations between successive images. ORL face database contains 10 different images for each of the 40 distinctive subjects. Subjects are imaged at different times, with varying lighting conditions, facial expressions and facial details.

Image databases are pre-cropped and resized with a 2 fold dimension reduction. Each database is randomly divided into testing and training image sets, which are decomposed using curvelet transform at 3 scales and 8 orientations. Approximate curvelet coefficients are dimensionally reduced using B2DPCA, vectorized, trained and tested using ELM. Fast learning and testing speed offered by ELM enabled us to repeat the experiments several times and calculate average results. We have compared our ELM based recognition scheme (50 hidden neurons) against PCA+LDA approach utilizing kNN of neighborhood size 5.

The recognition accuracy achieved for Sheffield and ORL databases using varying number of principal components is compared in Table I. The recognition achieved using our proposed method consistently outperforms PCA+LDA. The improvements in accuracy prove that our method is suitable to deal with challenging databases (views ranging from front to left and right). It is worth mentioning that increasing the number of principal components does not necessarily increase accuracy and the use of localized information for face recognition may be exploited to generate improved results.

TABLE I

RECOGNITION RATE(%): ORL AND SHEFFIELD DATABASE

Number of Components	ORL		Sheffield	
	[11]	Proposed	[11]	Proposed
5	75	99.44	93.89	93.99
10	88.13	96.33	96.11	99.31
15	90	99.61	97.78	99.8
20	93.12	99.93	99.44	99.91
25	92.5	99.93	99.44	100
30	94.3	99.96	98.88	100
35	96	99.99	98.46	100
40	94.37	100	97.12	100
45	93.75	100	97.77	100
50	93.75	100	97.22	100

VI. CONCLUSION

In this paper an appearance-based face recognition scheme with manual face localization is proposed. Curvelet transform was used to transform the image into a new domain and to calculate initial feature vectors. The feature vectors were dimensionally reduced using B2DPCA and classified using ELM. Remarkable improvements in accuracy were achieved in comparison to PCA+LDA based approach. Authors believe that complimentary localized feature information can further improve precision.

REFERENCES

[1] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma and N. Otsu, *Face Recognition System using Local Autocorrelations and Multiscale Integration*. IEEE Trans. on PAMI 18(10), pp. 1024-1028, 1996.

[2] R. Brunelli and T. Poggio, *Face Recognition: Features vs Templates*. IEEE Trans. PAMI 15(10), pp. 1042-1053, 1993.

[3] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, *Face recognition: A Literature Survey*. ACM Computing Surveys, 35(4), pp. 399-458, 2003.

[4] M. Kirby and L. Sirovich, *Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces*. IEEE Trans. on PAMI, 12(1), pp. 103-108, 1990.

[5] M. S. Bartlett, J. R. Movellan and T. J. Sejnowski, *Face Recognition by Independent Component Analysis*, IEEE Trans. on Neural Networks, 13(6), pp. 1450-1464, 2002.

[6] J. Lu, K. N. Plataniotis and A. N. Venetsanopoulos, *Face Recognition using LDA-Based Algorithms*. IEEE Trans. on Neural Networks, 14(1), pp. 195-200, 2003.

[7] G. C. Feng, P. C. Yuen and D. Q. Dai, *Human Face Recognition using PCA on Wavelet Subband*. Journal of Electronic Imaging, 9(2), pp. 226-233, 2000.

[8] J. T. Chien and C. C. Wu, *Discriminant Wavelet-Faces and Nearest Feature Classifiers for Face Recognition*. IEEE Trans. on PAMI, 24(2), pp. 1644-1649, 2002.

[9] B. L. Zhang, H. Zhang and S. Sam Ge, *Face Recognition by Applying Wavelet Subband Representation and Kernel Associative Memory*. IEEE Trans. on Neural Networks, 15(1), pp. 166-177, 2004.

[10] T. Mandal and Q. M. Jonathan Wu, *Face Recognition using Curvelet Based PCA*. Proceedings of IEEE ICPR, pp. 1-4, 2008.

[11] T. Mandal, Q. M. Jonathan Wu and Y. Yuan, *Curvelet Based Face Recognition via Dimension Reduction*. Elsevier Signal Processing, 89(3), pp. 2345-2353, 2009.

[12] J. L. Starck, *Image Processing by the Curvelet Transform*. Powerpoint presentation.

[13] E. J. Cands and D. L. Donoho, *Curvelets - A Surprisingly Effective Nonadaptive Representation for Objects with Edges*. Curves and Surface Fitting: Vanderbilt University Press, Nashville, pp. 105-120, 2000.

[14] E. J. Candes, L. Demanet, D. L. Donoho and L. Ying, *Fast Discrete Curvelet Transforms*. Multiscale Model Simulation, 5(3), pp. 861-899, 2006.

[15] L. Wiskott, J. M. Fellus, N. Kruger and C. VonDerMalsburg, *Face Recognition by Elastic Bunch Graph Matching*. IEEE Trans. on PAMI, 19(7), pp. 775-779, 1997.

[16] J. Yang, D. Zhang and A.F. Frangi, J. Yang, *Two-dimensional PCA: A new approach to appearance based face representation and recognition*. IEEE Trans. on PAMI, 26(1), pp. 131-137, 2004.

[17] G. Huang, Q. Zhu and C. Siew, *Extreme learning machine: Theory and applications*. Elsevier Neurocomputing, 70, pp. 489-501, 2006.

[18] D. B. Graham and N. M. Allinson, *Characterizing virtual Eigensignatures for general purpose face recognition*. In Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences, 163, pp. 446-456, 1998.

[19] F. Samaria and A. Harter, *Parameterisation of a stochastic model for human face identification*. 2nd IEEE Workshop on Applications of Computer Vision, pp. 138-142, 1994.