A GENERIC FACE REPRESENTATION APPROACH FOR LOCAL APPEARANCE BASED FACE VERIFICATION

Hazim Kemal Ekenel, Rainer Stiefelhagen
Interactive Systems Labs, Universität Karlsruhe (TH)
76131 Karlsruhe, Germany
{ekenel, stiefel}@ira.uka.de
web: http://isl.ira.uka.de/face_recognition/

Abstract

In this paper we present the experimental results of a generic local appearance based face representation approach obtained from the first and fourth experiments of the Face Recognition Grand Challenge (FRGC) version 1 data. The introduced representation approach is compared with the baseline system with the standard distance metrics of L1 norm, L2 norm and cosine angle. The experimental results show that the proposed local appearance based approach provides better and more stable results than the baseline system -holistic Eigenfaces- approach.

1.Introduction

Since 1990s, with the introduction of Eigenfaces approach [1] and with the comparative study [2] favouring the appearance based template matching approach over the geometric, feature based approach, appearance based methods have dominated the face recognition research. One very intriguing observation has been derived from over ten years of appearance based face recognition research: Although local appearance information –using local regions of salient features- has been shown to be superior to the holistic information -using whole face template- [2,3], interestingly face recognition research has been focused on holistic approaches and local appearance based face recognition has been ignored in a great extent. It has not had as much impact as the holistic approach, and compared to the plethora of the holistic methods, only a few techniques have been proposed to perform local appearance based face recognition. The main reason for this is that the initial local appearance based approaches [2,3] require the detection of salient features –i.e. eyes- which may not be an easy task. Moreover, erroneous detection of these local regions may lead to severe performance drops. Recently, a more generic local appearance based approach has been proposed, that divides the input face image into non-overlapping blocks to perform Eigenfaces locally on each block [4]. The experiments conducted in this study showed that the proposed method outperforms the standard holistic Eigenfaces approach under variations of expression and illumination.

In [5] discrete cosine transform (DCT) is utilized for local appearance based face recognition. In this study, the input face image is partitioned into 8x8 pixel blocks, and on each block DCT is performed. The most relevant DCT features are extracted using zig-zag scan and the obtained features are fused either at the feature level or at the decision level for face recognition. The approach is extensively tested on the CMU PIE [6] and Yale [7] face databases. It is compared with the well known holistic approaches –Eigenfaces [1], Fisherfaces [7], two face recognition architectures of independent component analysis (ICA) [8]- and with the other local appearance based method that uses principal component analysis (PCA) –Modular PCA- [4] to extract features from each local region. The experimental results show that the proposed local appearance based approach performs significantly better than the holistic approaches. It also outperforms modular PCA approach [4] which indicates that DCT is a better choice than PCA for local appearance based face representation. For detailed experimental results please see [5]. Besides the performance improvement, the proposed approach has the advantages of using a data independent basis and fast computation of the DCT features. Using a data independent basis implies that there is no need to build a data specific space as in the Eigenfaces approach. Moreover, it is known that face recognition algorithms' performance deteriorates when they are performed on compressed images [9], i.e. jpeg compressed. Since the proposed algorithm just follows the same processing steps in jpeg compression standard (except the quantization step) – performing DCT on 8x8 pixels image blocks and extracting coefficients via zig-zag scan- for feature extraction, it's less effected from this problem.

In this study, following the idea presented in [5] and being encouraged with the results obtained under the recognition task, we tested our generic local appearance based face representation approach under the verification task using the Face Recognition Grand Challenge (FRGC) evaluation data version 1. To have a proper comparison of the representation steps we used the pre-processed input images of the baseline system and the standard distance metrics used in face recognition/verification research, namely L1 norm, L2 norm and cosine angle. We haven't addressed any distance/score normalization issues related to the verification task. Our goal is to introduce a generic representation approach that can be accepted as a baseline for local appearance based face recognition/verification.

The organization of the paper is as follows. In Section 2, discrete cosine transform, the way it has been utilized in the face recognition research and the way we use it as a feature extraction/representation step is explained. Obtained performance results on the first and fourth experiment sets of the Face Recognition Grand Challenge (FRGC) version 1 data are presented in Section 3. Finally, Section 4 concludes the paper.

2. Methodology

2.1. Discrete Cosine Transform

Discrete cosine transform (DCT) is a well-known signal processing tool widely used in compression standards due to its compact representation power. The 2-D discrete cosine transform of an NxN image is defined as:

$$C(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (1)$$

for $u, v = 0, 1, \dots, N-1$ where

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & \textit{for } u = 0 \\ \sqrt{\frac{2}{N}} & \textit{for } u = 1,2,\dots,N-1 \end{cases}$$

and the 2-D inverse discrete cosine transform is defined as

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \alpha(u)\alpha(v)C(u,v)\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right]$$

Obtained DCT basis functions for N = 4 can be seen in Figure 1 (each base is scaled separately for illustration purposes). As can be seen from the top-left part of the basis functions and also from equation (1), the (0,0) component represents the average intensity value of the image, which can be directly effected by illumination variations.

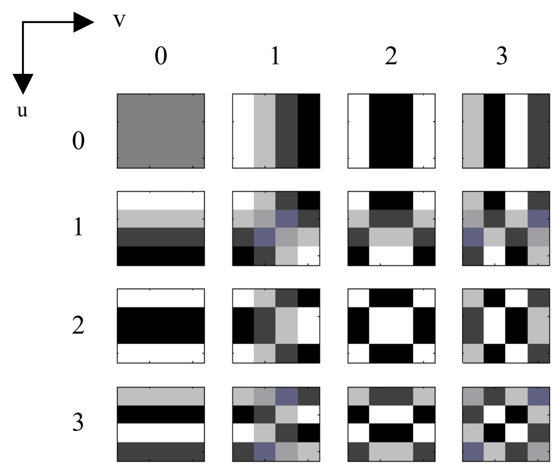

Figure 1. DCT basis functions for N = 4

2.2. Discrete Cosine Transform in Face Recognition Research

Discrete cosine transform has been used as a feature extraction step in various studies on face recognition. In [10] the DCT coefficients of the entire image or its blocks are computed and from the obtained coefficients only a subset of them are selected by diagonally scanning the upper-left part. The coefficients are then given as an input to a multi-layer perceptron. In [11], DCT is performed on the entire image and a square subset (i.e. 7x7) of the DCT coefficients from the top-left part is used as the feature vector. A derived coefficient set, called mod 2 feature set, from the DCT coefficients are proposed in [12] for face-based identity verification. 8x8 pixels blocks are used having 50% overlap between horizontally and vertically neighboring blocks. After ordering the DCT coefficients according to zig-zag scan the first three coefficients are replaced with their horizontal and vertical deltas which are suggested to represent transitional spatial information. Gaussian mixture models (GMM) are used for modelling the distribution of extracted feature vectors and verification is done comparing the average log-likelihood value of the claimant being the genuine and the average log-likelihood value of the claimant being the impostor. In [13], network of networks (NoN) model is fed by DCT coefficients. In [14] DCT coefficients are used to represent the image blocks in a compact form for embedded HMM based classification.

2.3. Local Appearance Based Face Representation using Discrete Cosine Transform

As presented in Section 2.2, up to now, discrete cosine transform has been performed either in a holistic appearance-based sense [12], or in a local appearance-based sense ignoring the spatial information in some extent during the classification step by feeding some kinds of neural networks with local DCT coefficients or by modelling them with some kinds of statistical tools [11,13,14,15].

In the proposed representation scheme DCT is used for local appearance modelling. It preserves spatial information for classification purposes. Furthermore, it follows the processing steps of jpeg compression standard which makes the proposed approach a more generic, more easily applicable and widely acceptable method. Local appearance based face representation can be performed in the following way. First, the detected and aligned input face image is partitioned into 8x8 pixels blocks. On each block DCT is performed and the DCT coefficients are extracted. The obtained DCT coefficients are ordered using zig-zag scanning. The first coefficient is removed since it only represents the average intensity value of the block and from the remaining coefficients the first M of them are selected resulting an M-dimensional local feature vector. Finally, the DCT coefficients extracted from each block are concatenated to construct the feature vector. As in [5], the local observations can be also fused at the decision level. However, it requires a more elaborate effort to perform decision fusion. Hence, for the sake of keeping the approach simple and generic we only performed fusion at the feature level. Besides, experimental results show that feature fusion performs better than decision fusion when the number of individuals to be identified increases.

3. Experimental Results

We conducted Experiment 1 and 4 on the Face Recognition Grand Challenge (FRGC) evaluation data version 1. We used the pre-processed images of the baseline system and we scaled them to 64x64 pixels resolution. We used three standard distance metrics for verification, namely the L1 norm, the L2 norm and the cosine angle which corresponds to the COV distance metric in the Biometric Experimentation Environment (BEE).

3.1. Experimental results on Experiment 1 of FRGC version 1 data set

The first experiment of FRGC version 1 data set corresponds to the controlled indoor still versus controlled indoor still matching scenario. The target and query sets contain 943 face images. Both the target and query sets consist of the same face image data, therefore after the verification process the diagonal elements of the obtained similarity matrix should be masked out.

The proposed approach is tested with different local feature dimensions and compared with the baseline Eigenfaces approach. The number of eigenvectors used in the Eigenfaces approach is 109, which is the default number in BEE. The equal error rates (EER) and the verification rates at 0.1% false acceptance rate, obtained with varying local feature dimensions can be seen from Figures 2 and 3. As can be observed the performance converges very quickly, even using 2 DCT coefficients per block suffices to obtain satisfying results. When the dimension becomes too high, the performance drops slightly. Another interesting observation from these figures is that there are no significant performance differences between different distance metrics. This is a desirable property, it implies that the choice of the distance metric would not effect the system much and the results are guaranteed to be stable. On the other hand, it can be observed from the ROC curves –Figures 4,5,6- and Tables 1,2, that the baseline Eigenfaces approach is very sensitive to the distance metric. The equal error rates and verification rates vary very much with different distance metrics. Overall, the proposed approach outperforms the baseline Eigenfaces approach significantly when the L2 norm or cosine angle metrics are used. With L1 norm, it performs slightly better when false accept rate exceeds ~10%.

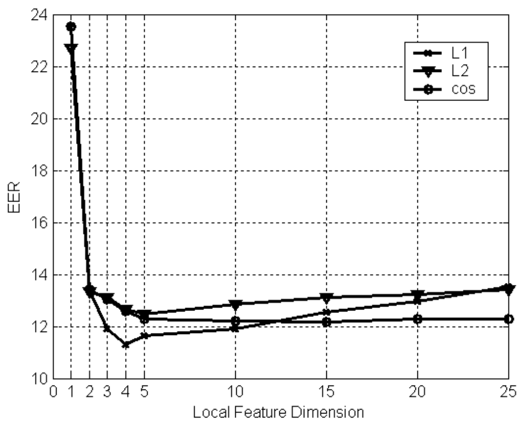

Figure 2. Equal Error Rate versus Local Feature Dimension

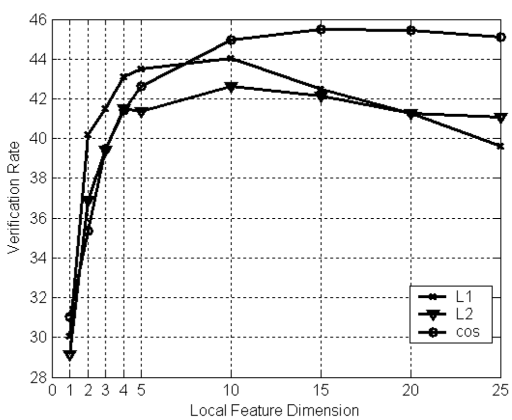

Figure 3. Verification Rate @ 0.1% False Accept versus Local Feature Dimension

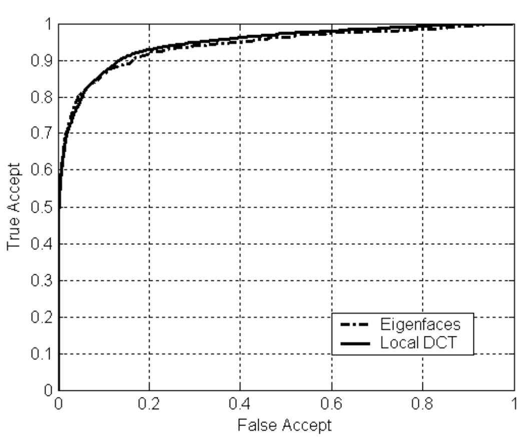

Figure 4. ROC curve when the distance metric is L1 norm –5 coefficients per block in Local DCT-

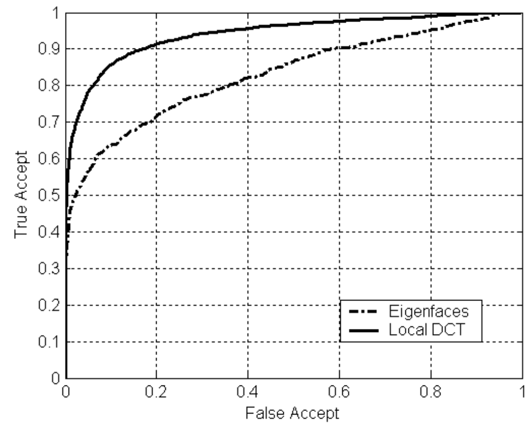

Figure 5. ROC curve when the distance metric is L2 norm –5 coefficients per block in Local DCT-

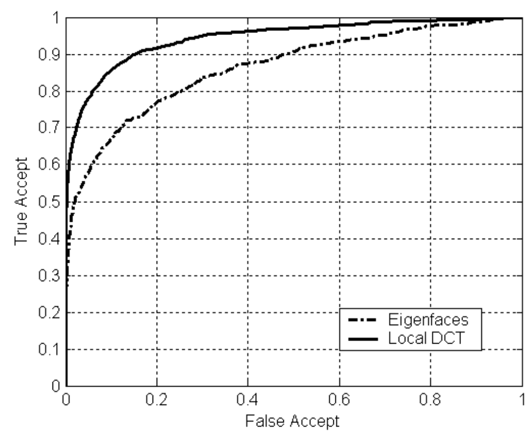

Figure 6. ROC curve when the distance metric is cosine angle –5 coefficients per block in Local DCT-

	Eigenfaces	**Local DCT**
L1	12.07%	11.64%
L2	24.91%	12.47%
cos	21.96%	12.29%

Table 1. Equal Error Rates of Eigenfaces and Local DCT –5 coefficients per block in Local DCT-

	Eigenfaces	Local DCT
L1	43.91%	43.48%
L2	26.97%	41.35%
cos	21.22%	42.64%

Table 2. Verification rate of Eigenfaces and Local DCT @ 0.1% false accept rate –5 coefficients per block in Local DCT-

3.2. Experimental results on Experiment 4 of FRGC version 1 data set

The fourth experiment of FRGC version 1 data set corresponds to the controlled indoor still versus uncontrolled still matching scenario. The target and query sets contain 943 face images. The target images are the “controlled indoor stills” and the query images are the “uncontrolled stills”.

Again, the proposed approach is tested with different local feature dimensions and compared with the baseline Eigenfaces approach. The number of eigenvectors used in the Eigenfaces approach is as in the first experiment, 109, which is the default number in BEE. We observed a similar outcome as in Experiment 1: the performance converges very quickly; when the dimension becomes too high, the performance drops slightly; no significant performance differences between different distance metrics. Overall, the proposed approach outperforms the baseline Eigenfaces approach significantly when the L2 norm or cosine angle metrics are used. With L1 norm, it performs slightly better when false accept rate exceeds ~25%. We didn’t present the verification rates, since they were very low. The baseline systems’ verification rates at 0.1% false accept rate were around 1% when L2 norm or cosine angle is used as distance metric. With L1 norm, it was around 5%. On the other hand the proposed approach’s verification rates were around 4%.

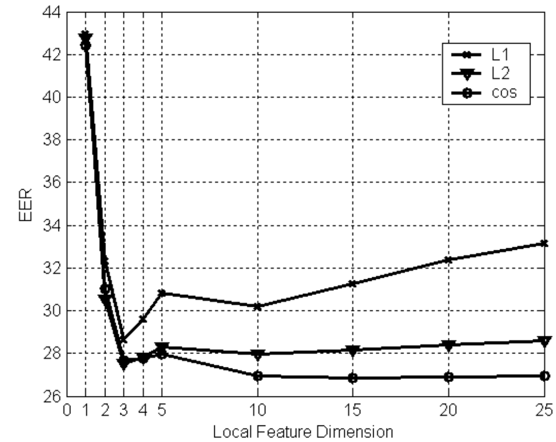

Figure 7. Equal Error Rate versus Local Feature Dimension

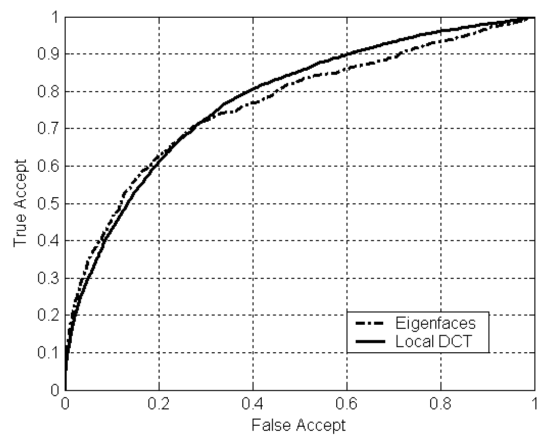

Figure 8. ROC curve when the distance metric is L1 norm –3 coefficients per block in Local DCT-

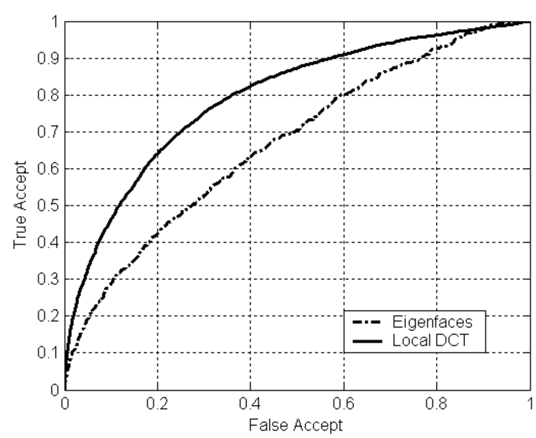

Figure 9. ROC curve when the distance metric is L2 norm –3 coefficients per block in Local DCT-

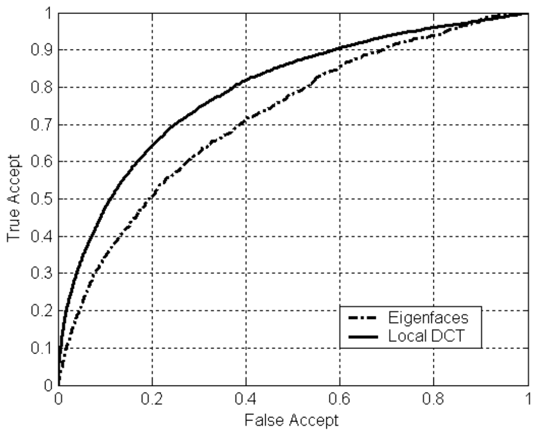

Figure 10. ROC curve when the distance metric is cosine angle –3 coefficients per block in Local DCT-

	Eigenfaces	Local DCT
L1	28.72%	28.64%
L2	38.54%	27.52%
cos	34.17%	27.65%

Table 3. Equal Error Rates of Eigenfaces and Local DCT –3 coefficients per block in Local DCT-

4. Conclusions

In this paper we presented and discussed the experimental results of a generic local appearance based face representation approach obtained from the first and fourth experiments of the Face Recognition Grand Challenge (FRGC) version 1 data. The introduced method is proposed as a baseline for local appearance based face recognition/verification. The approach is compared with the baseline –holistic Eigenfaces- system with the standard distance metrics of L1 norm, L2 norm and cosine angle. The obtained results show that the proposed generic local appearance based approach provides better and more stable results than the baseline Eigenfaces approach. The performance improvement is significant when the L2 norm or cosine angle is used. With L1 only a slight improvement is obtained. Besides the improvements in the verification rates, the consistency of the proposed approach's performance over different metrics makes it a more favourable face verification method.

5. References

[1] M. Turk and A. Pentland, "Eigenfaces for recognition", Journal of Cognitive Science, pp. 71–86, 1991.
[2] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates", IEEE Trans. on PAMI, Vol. 15, No. 10, pp. 1042-1052, 1993.
[3] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition", In Proc. of CVPR'94, 1994.
[4] R. Gottumukkal and V.K. Asari, "An improved face recognition technique based on modular PCA approach", Pattern Recognition Letters, 25(4), March 2004.
[5] H.K. Ekenel, R. Stiefelhagen, "Local Appearance based Face Recognition Using Discrete Cosine Transform", 13th European Signal Processing Conference (EUSIPCO 2005), Antalya, Turkey, September 2005.
[6] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database", In Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition, 2002.
[7] J.H.P.N. Belhumeur and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):711–720, 1997.
[8] M. Bartlett, J. Movellan, and T. Sejnowski, "Face recognition by independent component analysis", IEEE Transactions on Neural Networks, 13(6):1450–1454, 2002.
[9] M. Kamasak, B. Sankur, "Face Recognition under Lossy Compression", PRIP'99: Pattern Recognition and Information Processing, 27-32, Minsk, Belarus, May 1999.
[10] Z. Pan and H. Bolouri, "High speed face recognition based on discrete cosine transforms and neural networks", Technical report, University of Hertfordshire, UK, 1999.
[11] Z. M. Hafed and M. D. Levine, "Face recognition using the discrete cosine transform", International Journal of Computer Vision, 43(3), 2001.
[12] C. Sanderson and K. K. Paliwal, "Features for robust face-based identity verification", Signal Processing, 83(5), 2003.
[13] W. L. Scott, "Block-level Discrete Cosine Transform Coefficients for Autonomic Face Recognition", PhD thesis, Louisiana State University, USA, May 2003.
[14] A. Nefian, "A Hidden Markov Model-based Approach for Face Detection and Recognition", PhD thesis, Georgia Institute of Technology, 1999.