LOCAL APPEARANCE BASED FACE RECOGNITION USING DISCRETE COSINE TRANSFORM

Hazim Kemal Ekenel, Rainer Stiefelhagen

Interactive Systems Labs, University of Karlsruhe
Am Fasanengarten 5, 76131, Karlsruhe, Germany
phone: + (49) 721 6083564, fax: + (49) 721 607721, email: {ekenel, stiefel}@ira.uka.de
web: http://isl.ira.uka.de/face_recognition

Abstract

In this paper, a local appearance based face recognition algorithm is proposed. In the proposed algorithm local information is extracted using block-based discrete cosine transform. Obtained local features are combined both at the feature level and at the decision level. The performance of the proposed algorithm is tested on the Yale and CMU PIE face databases, and the obtained results show significant improvement over the holistic approaches.

1. INTRODUCTION

Face recognition, as an unsolved problem under the conditions of pose and illumination variations [17], still attracts significant research efforts. The main reasons for the ongoing research are: the increased need for natural identification for authentication in the networked society, for surveillance, for perceptual user interfaces, and the lack of robust features and classification schemes for the face recognition task.

Face recognition algorithms evolved from anthropometrical feature-based approaches in the 70's [6] to appearance-based holistic approaches in the 90's [15, 2]. In addition to these techniques, local appearance based approaches have proved to be promising. Elastic bunch graph matching (EBGM) [16], where local information is derived using Gabor wavelets, is one of the best performing algorithms in the FERET evaluations [11], and modular eigenspaces, where only eye and nose regions are used for identification, is shown to be superior to the holistic eigenface approach [10]. Moreover, it's known that variations on the facial appearance caused by i.e. occlusion, illumination and expression can lead to modifications on the entire representation coefficients in a holistic representation scheme. In this respect the approach of analyzing faces locally is believed to perform superior to the holistic appearance-based approaches, where a local change effects only the corresponding part of the representation and does not modify the representation vector as a whole.

The main goal of this work is to show that local appearance based face recognition is more robust against variations on facial appearance than the traditional holistic approaches (PCA, LDA, ICA). In this paper we utilize local information by using block-based discrete-cosine transform (DCT). The main idea is to mitigate the effects of expression, illumination and occlusion variations by performing local analysis and by fusing the outputs of extracted local features at the feature and at the decision level. The reason for preferring DCT over Karhunen-Loeve transform (KLT), which is known to be the optimal transform in terms of compactness of representation, is mainly because of its data independent bases. To construct the appropriate bases by KLT for data representation, one has to align training face images properly, otherwise the basis images can have noisy appearance. The effect of alignment can be seen in Fig. 1, where the first row corresponds to the obtained basis images by performing KLT on slightly misaligned training images from the CMU PIE database, and the second row corresponds to the obtained basis images again by performing KLT but this time on properly aligned training images from the same database. Although alignment can be done for the entire face with respect to some facial landmarks such as the centers of the eyes, it is almost impossible to align local parts of the face as successful as the entire face image. Suitable landmarks for each part of the face cannot be easily found. Hence noisy basis images from the KLT on a training set of local parts are inevitable. Moreover, since DCT closely approximates KLT in the sense of information packing, it's a very suitable alternative for compact data representation.

Figure 1: Eigenface bases computed from mis-aligned (top) and well-aligned (bottom) images.

DCT has been used as a feature extraction step in various studies on face recognition. Up to now, either DCT features have been used in a holistic appearance-based sense [4], or local appearance-based sense which ignores spatial information during the classification step. In [9], the DCT coefficients obtained from the image blocks are given as an input to a multi-layer perceptron, in [12] local DCT coefficients are modelled with GMM, in [13] a network of networks (NoN) model is fed by DCT coefficients and finally in [8] they are used to represent the image block in a compact form for embedded HMM based classification scheme. Besides DCT based studies, in [3] KLT is performed on face images that are divided into smaller sub-images.

Here, we present a novel local appearance based face recognition approach, which is based on the well-known DCT for local representation, and which preserves spatial information. Moreover, we discuss the important problem of fusing the local observations, and we investigate fusion methods both at the feature level and at the decision level.

The remainder of the paper is organized as follows. In Section 2, discrete cosine transform and fusion schemes used in the study are explained. Experimental results are presented and discussed in Section 3. Finally, in Section 4, conclusions and future recommendations are given.

2. METHODOLOGY

2.1 Discrete Cosine Transformation (DCT)

DCT is a well-known signal analysis tool used in compression standards due to its compact representation power. Although Karhunen-Loeve transform (KLT) is known to be the optimal transform in terms of information packing, its data dependent nature makes it unfeasible for use in some practical tasks. Furthermore, DCT closely approximates the compact representation ability of the KLT, which makes it a very useful tool for signal representation both in terms of information packing and in terms of computational complexity due to its data independent nature.

2.2 Local Appearance Based Face Representation

Local appearance based face representation is a generic local approach and does not require detection of any salient local regions, such as eyes, as in the modular or component based approaches [5, 10] for face representation. Local appearance based face representation can be performed as follows: A detected and normalized face image is divided into blocks of 8x8 pixels size. Each block is then represented by its DCT coefficients. The reason for choosing a block size of 8x8 pixels is to have small-enough blocks in which stationarity is provided and transform complexity is kept simple on one hand, and to have big enough blocks to provide sufficient compression on the other hand. The top-left DCT coefficient is removed from the representation since it only represents the average intensity value of the block. From the remaining DCT coefficients the ones containing the highest information are extracted via zig-zag scan.

2.3 Fusion

To fuse the local information, the extracted features from 8x8 pixels blocks can be combined at the feature level or at the decision level.

2.3.1 Feature Fusion

In feature fusion, the DCT coefficients obtained from each block are concatenated to construct the feature vector which is used by the classifier.

2.3.2 Decision Fusion

In decision fusion, classification is done separately on each block and later, the individual classification results are combined. To combine the individual classification results, we used the sum rule [7].

3. EXPERIMENTS

We compare the proposed local appearance-based approach with several well-known holistic face recognition approaches – Principal Component Analysis (PCA) [15], Linear Discriminant Analysis (LDA) [2], Independent Component Analysis (ICA) [1], global DCT [4] – as well as another DCT based local approach, which uses Gaussian mixture models for modeling the distributions of feature vectors [12]. This approach will be named "local DCT + GMM" in the remainder of the paper. Moreover, we also test a local appearance-based approach using PCA for the representation instead of DCT which will be named Local PCA in the paper. In all our experiments, except for the DCT+GMM approach, where the classification is done with Maximum-Likelihood, we use the nearest neighbor classifier with the normalized correlation d as the distance metric:

$$d = \frac{f_{training} \cdot f_{test}}{\|f_{training}\| * \|f_{test}\|} \quad (1)$$

Figure 2: Samples from the Yale database. First row: Samples from training set. Second row: Samples from test set.

3.1 Experiments on the Yale database

The Yale face database [2] consists of 15 individuals, where for each individual, there are 11 face images containing variations in illumination and facial expression. From these 11 face images, we use 5 for training, the ones with annotations "center light", "no glasses", "normal", "sleepy" and "wink". The remaining 6 images - "glasses", "happy", "left light", "right light", "sad", "surprised" - are used for testing. The test images with illumination from sides and with glasses are put in the test set on purpose in order to harden the testing conditions. The face images are closely cropped and scaled to 64x64 resolution. Fig. 2 depicts some sample images from the training and testing set.

In the first experiment, the performances of PCA, global DCT, local DCT and local PCA with feature fusion are examined with varying feature vector dimensions. Fig. 3 plots the obtained recognition results for the four approaches for varying number of coefficients (holistic and local approaches are plotted in different figures due to the difference in the dimension of used feature vectors in the classification). It can be observed that while there's no significant performance difference between PCA, local PCA and global DCT, local DCT with feature fusion outperforms these three approaches significantly. Fig. 3 shows that Local DCT outperforms Local PCA significantly at each feature vector dimension which indicates that using DCT for local appearance representation is a better choice than using PCA. Next, the block-based DCT with decision fusion is examined, again with varying feature vector dimensions. Table 1 depicts the obtained results. It can be seen that further improvement is gained via decision fusion. Using 20 DCT coefficients, 99% accuracy is achieved. For comparison, the results obtained when using PCA for local representation are also depicted in Table 1. Overall, the results obtained with PCA for local appearance

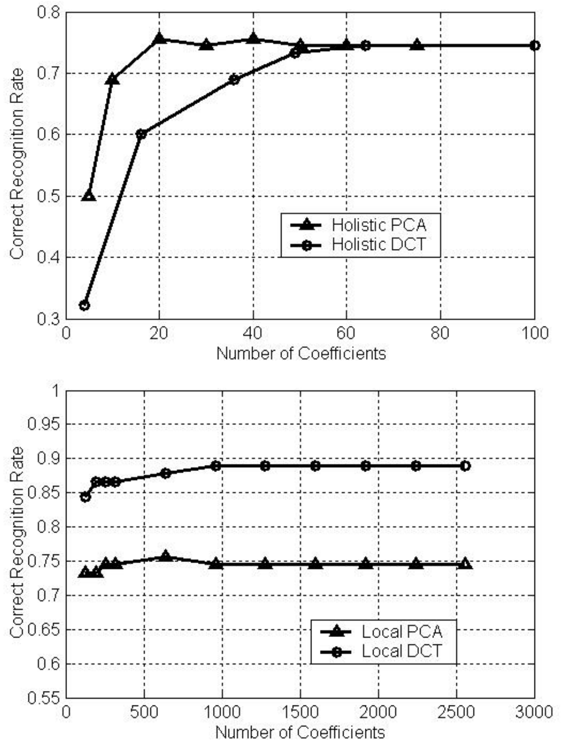

Figure 3: Correct recognition rate versus number of used coefficients on the Yale database. First row: Holistic PCA vs. holistic DCT. Second row: Local PCA vs. local DCT.

# coeff.	DCT	PCA
2	72.2%	57.8%
5	94.4%	70.0%
10	98.9%	71.1%
20	98.9%	72.2%

Table 1: Decision fusion results on the Yale database.

represenation are much lower than those obtained with the local DCT representation.

Table 2 compares the proposed local appearance-based approaches with the holistic approaches – PCA, LDA, two variants of ICA (ICA 1, ICA 2), global DCT and the local DCT+GMM method. In ICA1 and ICA2, as recommended in [1], most of the energy content is conserved during the prior PCA-stage. 40 eigenvectors are chosen corresponding to 97.92% of the energy content. For LDA, again the initial dimension is first reduced using PCA to the size of $X-K$, where here X denotes the total number of samples and K denotes the number of classes. Afterwards, LDA is performed and the dimension is further reduced to $K-1$, i.e. 14 in this case. The parameters used in the local DCT+GMM approach are the same as used in [12].

From the results depicted in Table 2 it can be seen that the proposed approaches using local DCT features outperform the holistic approaches as well as the local DCT features modeled with a GMM, which ignores location information.

Method	Reco. Rate
PCA (20)	75.6%
LDA (14)	80.0%
ICA 1 (40)	77.8%
ICA 2 (40)	72.2%
Global DCT (64)	74.4%
Local DCT (18) + GMM (8) as in [12]	58.9%
Local DCT + Feature Fusion (192)	**86.7%**
Local DCT (10) + Decision Fusion (64)	**98.9%**

Table 2: Overall comparison of methods on the Yale database. In brackets, the number of coefficients is given.

Figure 4: Samples from the CMU PIE database. First row: Samples from training set. Second row: Samples from test set.

3.2 Experiments on the CMU PIE database

The face database derived from the CMU PIE face database [14] consists of 2720 face images of 68 individuals. Each individual has 40 face images. 20 images are chosen for training that correspond to normal appearance and expression variations (from the expression and talking set), and the rest, which are taken under different illumination conditions, are used for testing. The face images are aligned and scaled to a resolution of 64x64 pixels.

As on the Yale face database, the performances of PCA, global DCT, local PCA and local DCT with feature fusion are examined with varying feature vector dimensions (Fig. 5). In this experiment, again, local DCT with feature fusion outperforms PCA, the global DCT approach and the local PCA. It can be seen that on the CMU PIE database, the global DCT approach performs worse than PCA. This shows that global DCT is more sensitive to the illumination variations than PCA.

In Table 3, the decision fusion results are shown. For comparison, the results using the PCA represenation are also depicted. As observed on the Yale database, the results obtained using PCA are considerably lower than the results obtained with DCT representation.

Table 4 summarizes the results obtained with the different approaches on the CMU PIE database. As in the experiments with the Yale database, it can be seen that our proposed

#coeff.	DCT	PCA
10	68.5%	45.7%
20	71.8%	46.8%

Table 3: Decision fusion results on the CMU PIE database.

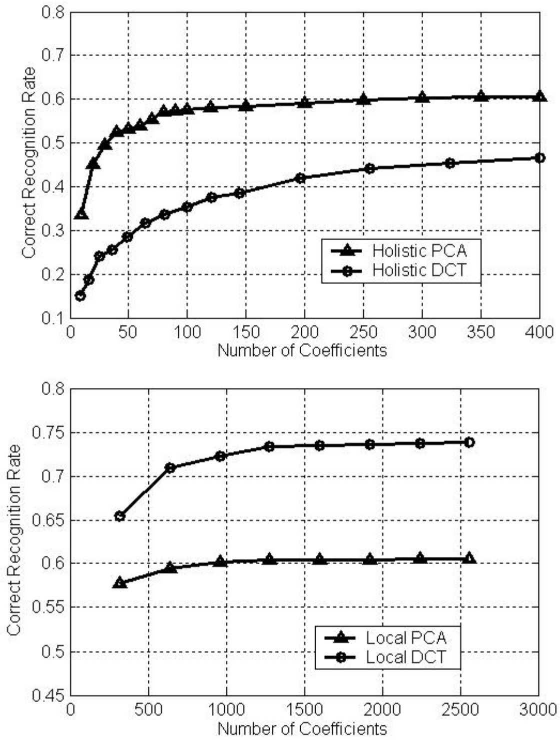

Figure 5: Correct recognition rate versus number of used coefficients on the CMU PIE database. First row: Holistic PCA vs. holistic DCT. Second row: Local PCA vs. local DCT.

local-appearance based approaches perform superior to the holistic approaches (PCA, LDA, ICA1, ICA2) as well as the local DCT+GMM approach. We think that the main reason for the poor performance of the local DCT+GMM method on this database is mainly the insufficient number of Gaussian mixtures to model the features.

By comparing the results obtained on the Yale database with the ones obtained on the CMU PIE database, it can be observed that the correct recognition rates obtained on the CMU PIE database are lower than the ones obtained on the Yale database. We think that there are two main reasons for this. The first one is, in the Yale database for testing we have face images containing expression variations as well as illumination variations (training face images contain only expression variations), whereas in CMU PIE all the testing data consists of face images under illumination variations (training face images contain only expression variations) which makes it a harder case. The other reason is the number of individuals in the databases (15 vs. 68).

4. SUMMARY AND CONCLUSIONS

In this paper we have presented a novel local appearance-based face recognition approach, which utilizes the block-based discrete cosine transform for local representation and which preserves the spatial information of the extracted DCT features. The proposed approach is quite generic and can be applied to any object classification problem, since it does not require any object-specific detection of salient parts.

We investigated fusion schemes, both at the feature level and at the decision level. We conducted extensive experiments on the Yale and the CMU PIE face databases using different feature vector dimensions and we thoroughly compared the proposed algorithm with well-known holistic appearance-based approaches (PCA, LDA, ICA1, ICA2) as well as with another local appearance based approach [12] and with local PCA [3].

Method	Reco. Rate
PCA (80)	57.1%
LDA (67)	59.5%
ICA 1 (200)	59.1%
ICA 2 (200)	51.8%
Global DCT (256)	44.1%
Local DCT (18) + GMM (8) as in [12]	12.0%
Local DCT + Feature Fusion (640)	**70.9%**
Local DCT (10) + Decision Fusion (64)	**68.5%**

Table 4: Overall comparison of methods on the CMU PIE database.

From the experimental results, it's apparent that the proposed local face recognition approach outperforms the holistic approaches, no matter whether the information is fused at the feature level or at the decision level. The experimental results also indicate that DCT is a better choice than PCA for local appearance based face representation. Furthermore, our experiments show that maintaining spatial information in the feature vector improves face recognition performance.

While, on the Yale face database decision fusion performs superior, on the CMU PIE database, feature fusion performs better. A possible reason for this difference could be the class size. As class size increases, it seems like it becomes harder to classify faces by local observations. To overcome this problem, we are planning to try a different approach for classification - a hierarchical classification scheme, in which, first, the number of potential candidates will be decreased using classification based on feature fusion, and then decision fusion will be performed for classification within the resulting subset from the first step. We are also planning to investigate a hybrid fusion approach, in which features obtained from the neighboring blocks are combined first, and then decision fusion is performed over the larger blocks.

5. ACKNOWLEDGEMENTS

This work was sponsored by the European Union under the integrated project CHIL, *Computers in the Human Interaction Loop*, contract number 506909.

REFERENCES

[1] M. S. Bartlett et al. Face recognition by independent component analysis. *IEEE Trans. on Neural Networks*, 13(6):1450–1454, 2002.

[2] P. N. Belhumeur et al. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on PAMI*, 19(7):711–720, 1997.

[3] R. Gottumukkal and V. K. Asari. An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters*, 25(4), 2004.

[4] Z. M. Hafed and M. D. Levine. Face recognition using the discrete cosine transform. *International Journal of Computer Vision*, 43(3), 2001.

[5] B. Heisele et al. Face recognition with support vector machines: Global versus component-based approach. In *ICCV*, pages 688–694, 2001.

[6] T. Kanade. Picture processing by computer complex and recognition of human faces. Technical report, Kyoto Univ., Dept. Inform. Sci., 1973.

[7] J. Kittler et al. On combining classifiers. *IEEE Trans. on PAMI*, 20(3), 1998.

[8] A. Nefian. *A Hidden Markov Model-based Approach for Face Detection and Recognition*. PhD thesis, Georgia Institute of Technology, 1999.

[9] Z. Pan and H. Bolouri. High speed face recognition based on discrete cosine transforms and neural networks. Technical report, Univ. of Hertfordshire, 1999.

[10] A. Pentland et al. View-based and modular eigenspaces for face recognition. In *CVPR'94*, 1994.

[11] P. J. Phillips et al. The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16(5), 1998.

[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] T. Sim et al. The CMU pose, illumination, and expression (PIE) database. In *5th Intl. Conf. on Automatic Face and Gesture Recognition*, 2002.

[15] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Science*, pages 71–86, 1991.

[16] L. Wiskott et al. Face recognition by elastic bunch graph matching. *IEEE Trans. on PAMI*, 19(7):775–780, 1997.

[17] W. Zhao et al. Face recognition: A literature survey. *ACM Computing Survey*, 35(4):399–458, 2003.