Analysis of Local Appearance-based Face Recognition: Effects of Feature Selection and Feature Normalization

Hazım Kemal Ekenel, Rainer Stiefelhagen
Interactive Systems Labs, Computer Science Department, Universität Karlsruhe (TH)
Am Fasanengarten 5, 76131, Karlsruhe, Germany
{ekenel,stiefel}@ira.uka.de

Abstract

In this paper, the effects of feature selection and feature normalization to the performance of a local appearance based face recognition scheme are presented. From the local features that are extracted using block-based discrete cosine transform, three feature sets are derived. These local feature vectors are normalized in two different ways; by making them unit norm and by dividing each coefficient to its standard deviation that is learned from the training set. The input test face images are then classified using four different distance measures: L1 norm, L2 norm, cosine angle and covariance between feature vectors. Extensive experiments have been conducted on the AR and CMU PIE face databases. The experimental results show the importance of using appropriate feature sets and doing normalization on the feature vector.

1. Introduction

Since the beginning of 1990s, appearance-based holistic approaches have been dominating the face recognition research [1-4]. Although local appearance information (using local regions of salient features) has been shown to be superior to the holistic information (using whole face template) in [2,5], 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 [5-9]. The main reason for this is that the initial local appearance based approaches [2,5] 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 in [6], that divides the input face image into non-overlapping blocks, without considering any salient region, to perform discrete cosine transform on each block. The experiments conducted in this study showed that the proposed method outperforms the well-known traditional holistic approaches [1,3,4] as well as a local appearance based approach which employs principal component analysis for local appearance representation [7].

Discrete cosine transform (DCT) has been used as a feature extraction step in various studies on face recognition [10-14]. Up to now, discrete cosine transform has been performed either in a holistic appearance-based sense [11], 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 [10,12,13,14]. On the other hand, in the proposed representation scheme in [6], spatial information is conserved, while using DCT for local appearance representation. The same approach is also tested for face verification [15] and for video-based face recognition [16]. In both of these studies, the proposed local appearance-based face recognition approach outperformed the holistic approaches.

In this paper, following the studies [6,15,16], the effects of feature selection and feature normalization to the performance of local appearance based face recognition scheme are investigated. Determining good feature sets, and feature normalization is specifically important for a DCT-based pattern recognition scheme. In terms of feature selection, in DCT, the first three coefficients correspond to specific information about the global statistics of the processed block. The first coefficient represents the average intensity value of the block, whereas the second and third coefficients represent the average vertical and horizontal intensity changes, respectively. Therefore, it is intriguing to observe the contributions of these features to the face recognition performance. Regarding feature normalization, there are two aspects. The first aspect is the total magnitude of each block's DCT coefficients. Since DCT is an orthogonal transformation and conserves all the energy of the processed input block, blocks with different brightness levels lead to DCT coefficients with different value levels. Because of this reason, it would be better

Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)
0-7695-2646-2/06 $20.00 © 2006 IEEE

when the extracted feature vectors' norm is normalized to a constant value. The other aspect is the value range of DCT coefficients. For example, the first coefficients have higher magnitudes than the later ones, thus having more influence on the classification results. To balance the contribution of each DCT coefficient, the coefficients are divided to their standard deviations that are learned from the training set. The effect of distance metrics on the performance is also analyzed. Four different distance metrics are used to classify face images: L1 norm, L2 norm, cosine angle and covariance between feature vectors.

The organization of the paper is as follows. In Section 2, discrete cosine transform is described briefly. Face recognition using block-based DCT is explained in Section 3. The used feature sets and the normalization techniques are introduced in Section 4. Experimental results are presented and discussed in Section 5. Finally, in Section 6, conclusions are given.

2. Discrete Cosine Transform

Discrete cosine transform (DCT) is a well-known signal analysis tool, widely used in feature extraction and compression applications 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 $u, v = 0, 1, \ldots, N - 1$ where

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

Obtained DCT basis functions for N = 4 can be seen in Fig. 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 Eq. 1, the (0,0) component represents the average intensity value of the image, which can be directly effected by illumination variations. From the figure, it can be also noticed that the (0,1) and (1,0) components represent the average vertical and horizontal intensity changes, respectively.

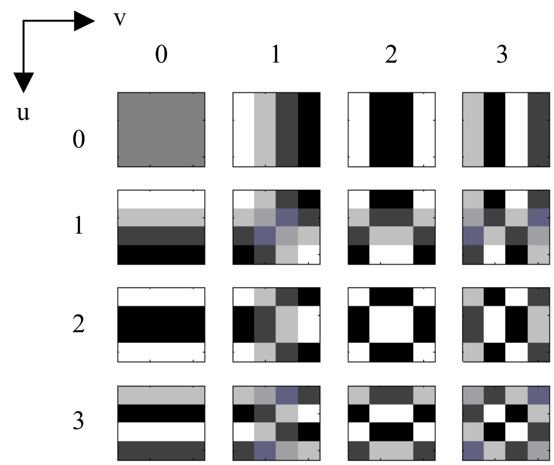

Figure 1. DCT basis functions for N = 4

3. Face Recognition Using DCT

In the local appearance-based face representation approach, a detected and normalized face image is divided into blocks of 8x8 pixels size. On each 8x8 pixels block, DCT is performed. The obtained DCT coefficients are ordered using zig-zag scanning (Fig. 2).

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

Figure 2. The order of DCT coefficients in zig-zag scan pattern

From the ordered coefficients, according to the feature selection strategy, *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 (Fig. 3).

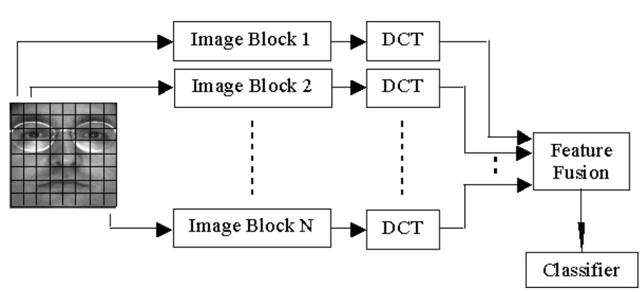

Figure 3. The diagram of feature extraction

Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)
0-7695-2646-2/06 $20.00 © 2006 IEEE

4. Feature Selection and Feature Normalization

Feature selection and feature normalization is specifically important for a DCT-based face recognition scheme. In this section, we introduce the used feature sets and the normalization techniques that we investigated.

4.1. Feature Selection

The first three DCT coefficients contain general information about the global statistics of the processed block of an image. While the first coefficient represents the average intensity value of the whole block, the second and third coefficients represent the average horizontal and vertical intensity change in the image block, respectively.

It is important to assess the contributions of these features to the recognition performance. Therefore, we investigated the use of three different feature sets for classification, where the local feature vectors are constructed by one of the following methods:

1. Selecting the first M DCT coefficients.
2. Removing the first coefficient, and selecting the first M DCT coefficients from the remaining ones.
3. Removing the first three coefficients, and selecting the first M DCT coefficients from the remaining ones.

We named the first feature set as DCT-all, the second one as DCT-0, and the third one as DCT-3.

4.2. Feature Normalization

There are two aspects in feature normalization. The first aspect is the total magnitude of each block's DCT coefficients. Since DCT is an orthogonal transformation and conserves all the energy of the processed input block, the blocks with different brightness levels lead to DCT coefficients with different value levels. Because of this reason, we normalized the local feature vector's, f's, magnitude to unit nom:

$$f_n = f / \|f\|, \tag{3}$$

where f_n represents the normalized feature vector.

The other aspect is the value range of DCT coefficients. For example, the first coefficients have higher magnitudes than the later ones (Fig. 4), thus having more influence on the classification results. To balance the contribution of each DCT coefficient; the coefficients, "f_i"s, are divided by their standard deviations, which are learned from a training set:

$$f_{n,i} = f_i / \sigma(f_i), \tag{4}$$

where $f_{n,i}$ represents the normalized ith coefficient and $\sigma(f_i)$ represents the standard deviation of the ith coefficient that is learned from the training set.

Note that, these presented normalization techniques are performed separately on the raw feature vector to obtain the normalized feature vector.

861	-158	28	-21	8	-2	-1	-1
87	-47	63	-68	18	0	-5	4
85	17	73	-19	-1	-2	-2	3
-1	-27	10	20	-10	15	-15	6
-7	6	-13	12	-12	15	-6	1
-10	16	-15	23	-15	12	-6	3
-7	10	-13	8	-7	9	1	3
-5	3	-7	6	-5	2	2	2

Figure 4. Sample transform output

5. Experiments

Two separate experiments are conducted to assess the effects of feature selection and feature normalization to the performance of the local appearance based face recognition scheme. The first experiment is conducted on the face images from the AR face database [17]. In this experiment, the main variation between the training and testing images is the time gap between the recording sessions. The second experiment is conducted on the face images from the CMU PIE database [18]. In this experiment, the main variation between the training and testing images is the illumination conditions. For comparison purposes, the same feature normalization techniques are also applied to the feature vectors extracted by principal component analysis (PCA) in the eigenfaces approach. Two feature sets are derived from the PCA coefficients, one, by selecting the first M PCA coefficients, and the other, by removing the first three coefficients – which are claimed to represent the illumination variations [3] – and then selecting the first M PCA coefficients from the remaining ones. We named the first feature set as PCA-all, and the second one as PCA-3.

For classification, the nearest neighborhood classifier is used in this study. Four different distance metrics are evaluated comparatively, namely, the L1 norm, the L2 norm, cosine angle, and covariance, defined as follows:

$$\text{L1:}\quad d = \sum_{m=1}^{M} \left| f_{training,m} - f_{test,m} \right| \tag{5}$$

$$\text{L2:}\ d = \left(\sum_{m=1}^{M} \left| f_{training,m} - f_{test,m} \right|^2 \right)^{1/2} \tag{6}$$

Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)
0-7695-2646-2/06 $20.00 © 2006 IEEE

$$\text{cos: } d = \frac{f_{training} * f_{test}}{\left\| f_{training} \right\| * \left\| f_{test} \right\|} \tag{7}$$

$$\text{cov: } d = \frac{\left(f_{training} - m_{training}\right) * \left(f_{test} - m_{test}\right)}{\left\| f_{training} - m_{training} \right\| * \left\| f_{test} - m_{test} \right\|} \tag{8}$$

where $f_{training,m}$ is the mth (m = 1,…, M) coefficient in the training feature vector $f_{training}$ and $m_{training}$ is the mean value of the training feature vector. Similarly, $f_{test,m}$ is the mth coefficient in the test feature vector f_{test} and m_{test} is the mean value of the test feature vector.

5.1. Experiments on the AR face database

The face database used in this experiment consists of 1540 face images of 110 individuals that are taken from the first and second sessions of the AR face database [17]. Each individual in the face database has 14 face images. Seven of these face images are from the first session and the remaining seven face images belong to the second session. Both of the images in separate sessions contain the same kind of variations that are annotated as "neutral expression", "smile", "anger", "scream", "left light on", "right light on", "all side lights on". The images from the first session are used for training and the ones from the second session are used for testing. The face images are aligned using the eye center locations and scaled to 64x64 pixels resolution. Sample images can be seen in Fig. 5.

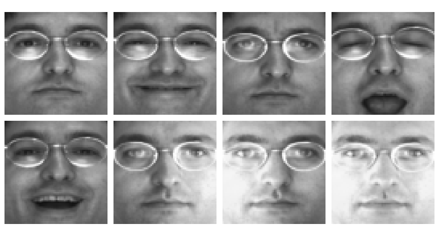

Figure 5. Sample face images from the AR face database. First row: Samples from the training set. Second row: Samples from the test set.

The correct classification rates of our local appearance-based face recognition scheme and the eigenfaces approach can be seen in Tables 1-3. The used feature vectors' dimension is 320, that is, 5 DCT coefficients per block in the local appearance-based approach and 320 PCA representation coefficients in the eigenfaces approach. In Table 1, the results obtained by using feature vectors without any normalization are shown. As can be observed, the best result is obtained by classifying the DCT-0 feature set using the L1 norm. The results derived by using feature vectors with unit norm are given in Table 2. Again, the L1 norm is found to be the best distance metric. Both the DCT-0 and DCT-3 feature sets provide the highest results. Finally, in Table 3, the results obtained by using feature vectors with normalized coefficients are given. In this experiment, DCT-all and DCT-0 features sets with L1 norm and PCA-all and PCA-3 feature sets with "cos" and "cov" distance metrics reached the best results.

Table 1. DCT and PCA scores on the AR database using feature vectors with no normalization

	DCT-all	DCT-0	DCT-3	PCA-all	PCA-3
L1	86.1%	**91.9%**	88.8%	87.2%	87.2%
L2	75.2%	82.6%	79.0%	77.0%	84.5%
Cos	82.0%	83.5%	80.6%	77.8%	86.9%
Cov	82.0%	83.5%	80.5%	77.8%	87.0%

Table 2. Local DCT and PCA scores on the AR database using unit norm feature vectors

	DCT-all	DCT-0	DCT-3	PCA-all	PCA-3
L1	91.0%	**94.8%**	**95.3%**	82.7%	89.9%
L2	80.3%	92.2%	93.9%	78.2%	86.9%
Cos	80.3%	92.2%	93.9%	77.8%	86.9%
Cov	80.0%	91.8%	93.9%	77.8%	87.0%

Table 3. DCT and PCA scores on the AR database using feature vectors with normalized coefficients

	DCT-all	DCT-0	DCT-3	PCA-all	PCA-3
L1	**93.8%**	**93.5%**	90.3%	79.6%	78.6%
L2	86.0%	85.8%	80.5%	78.2%	77.8%
Cos	88.1%	90.1%	88.6%	**93.6%**	**93.5%**
Cov	88.1%	89.9%	88.6%	**93.6%**	**93.5%**

To have an overall and thorough view, the best performing DCT, PCA feature set and distance metric couples from each normalization method are plotted for different feature vector dimensions in Fig. 6. Used local feature dimensions vary from two to ten. If there are more than one high performing couples as in Table 2 and Table 3, the ones that perform slightly better are chosen as the representative couples. In this figure's legend, "NN" corresponds to feature vectors with no normalization, "UN" corresponds to unit norm feature vectors, and "ON" corresponds to feature vectors with normalized coefficients. As can be seen the best performing combination is the unit norm DCT-3 feature set classified

IEEE COMPUTER SOCIETY

using L1 norm. If we look at the Table 2, we can also see that unit norm DCT-0 feature set with L1 norm also has very similar correct recognition rate.

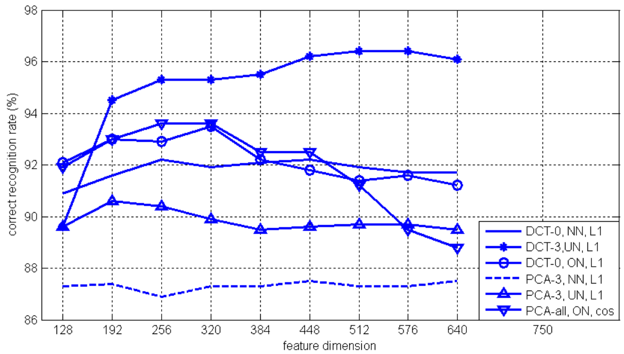

Figure 6. Correct recognition rate vs. feature dimension plot for feature selection, feature normalization and distance metric combinations

5.2. Experiments on the CMU PIE database

To observe the behavior of the feature selection and normalization techniques under a different variation between training and testing data, a second experiment is conducted using the CMU PIE database [18]. The face database used in this experiment consists of 1632 face images of 68 individuals that are taken from the "illumination" and "lighting" sets of the CMU PIE database [18]. Each individual in the face database has 24 face images –12 from the "illumination" set and 12 from the "lighting" set. The face images from the "lighting" set are used for training and the ones from the "illumination" set are used for testing. The face images are aligned using the eye center locations and scaled to 64x64 pixels resolution. Sample images can be seen in Fig. 7.

Figure 7. Sample face images from the CMU PIE database. First row: Samples from the training set. Second row: Samples from the test set.

The correct classification rates, on the CMU PIE database, of our local appearance-based face recognition scheme and the eigenfaces approach can be seen in Tables 4-6. The used feature vectors' dimension is again 320. In Table 4, the results obtained by using feature vectors without any normalization are shown. As can be observed, the best results are obtained by classifying the DCT-0 feature set with "cos" and "cov" distance metrics. The results derived by using unit norm feature vectors are given in Table 5. Here, the DCT-0 feature set with the L1 norm is found to be the best feature set-distance metric couple. Finally, in Table 6, the results obtained by using feature vectors with normalized coefficients are given. In this experiment, DCT-all and DCT-0 feature sets provided the best results using "cos" and "cov" distance metrics. Another interesting observation that can be derived from these experiments is the performance difference between PCA-all and PCA-3 feature sets. When there is no normalization on each individual coefficient, PCA-3 feature set performs significantly better than the PCA-all feature set (Tables 4-5). This finding supports the claim that the first three coefficients in PCA are effected from the illumination variations most.

Table 4. DCT and PCA scores on the CMU PIE database using feature vectors with no normalization

	DCT-all	**DCT-0**	**DCT-3**	**PCA-all**	**PCA-3**
L1	43.3%	83.9%	64.5%	67.0%	60.7%
L2	34.8%	73.9%	52.8%	36.2%	58.1%
Cos	82.4%	**88.0%**	78.1%	36.9%	79.0%
Cov	82.7%	**88.2%**	77.7%	37.1%	78.8%

Table 5. DCT and PCA scores on the CMU PIE database using unit norm feature vectors

	DCT-all	**DCT-0**	**DCT-3**	**PCA-all**	**PCA-3**
L1	92.3%	**97.4%**	93.5%	62.0%	78.7%
L2	89.7%	95.3%	90.1%	41.4%	80.6%
Cos	89.7%	95.3%	90.2%	36.9%	79.0%
Cov	90.0%	95.3%	90.3%	37.1%	78.8%

Table 6. DCT and PCA scores on the CMU PIE database using feature vectors with normalized coefficients

	DCT-all	**DCT-0**	**DCT-3**	**PCA-all**	**PCA-3**
L1	72.1%	84.2%	70.0%	45.3%	44.6%
L2	74.1%	77.6%	60.4%	41.1%	40.4%
Cos	**90.4%**	**91.9%**	84.6%	82.4%	82.7%
Cov	**90.7%**	**91.8%**	84.6%	82.1%	82.8%

As in Section 5.1, to have an overall and thorough view, the best performing DCT, PCA feature set and distance metric couples from each normalization method are plotted for different feature vector dimensions in Fig. 8. If there are more than one high performing couples as in Table 6, the ones that perform slightly better are chosen as the representative couples. In this figure's legend, "NN" corresponds to feature vectors with no normalization, "UN" corresponds to feature vectors with unit norm, and "ON" corresponds to feature vectors with normalized coefficients. As can be seen the best performing combination is the unit norm DCT-0 feature set with L1 norm. As can be observed from Fig. 8 (also from Fig. 6), when one uses higher dimensional feature vectors with normalized coefficients for classification, the performances decrease both in DCT and PCA. This shows that the coefficients that correspond to lower energy content does not contribute to the correct recognition performance as much as the ones that correspond to higher energy content. Therefore, when their ratio increases in the feature vector, the performance decreases.

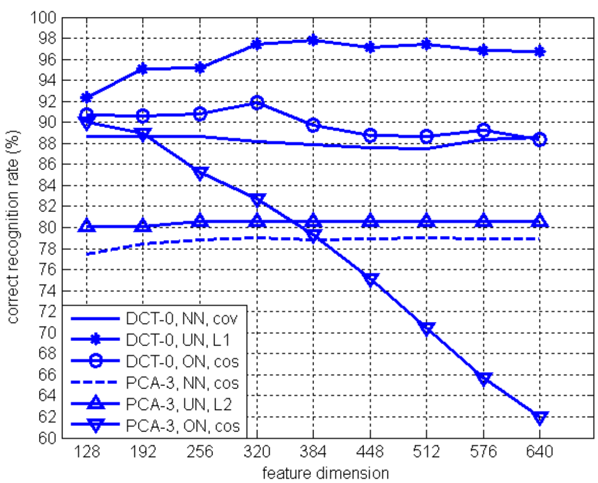

Figure 8. Correct recognition rate vs. feature dimension plot for feature selection, feature normalization and distance metric combinations

6. Conclusions

In this study, we investigated the effects of feature selection and feature normalization to the performance of a local appearance based face recognition scheme. We tested three different feature sets and two normalization techniques. We also analyzed the effects of distance metrics on the performance. Moreover, we applied feature selection and normalization to the holistic baseline approach – Eigenfaces – and conducted extensive comparative experiments. We found that the proposed unit norm DCT-0 feature set, in which the first DCT coefficient is removed from the feature set, together with using the L1 norm for classification, provides consistently high correct recognition rates. Similar results were also obtained from the experiments conducted on the FRGC version 2 database [19].

Table 7 and 8 give an overview of the improvements that are gained by using the proposed unit norm DCT-0 feature set together with the L1 norm, with respect to the best classification results without doing normalization. As one can observe, there is a significant decrease in the error rate, especially if the variation between training and testing is caused by illumination, as it is the case on the experiments performed on the CMU-PIE database.

Table 7. Improvement due to unit norm DCT-0 feature set with L1 norm on the AR face database

	Absolute Improvement in the Correct Recognition Rate (%)	Decrease in the Error Rate (%)
DCT-0	2.9	35.8
PCA-all	7.6	59.4
PCA-3	7.6	59.4

Table 8. Improvement due to unit norm DCT-0 feature set with L1 norm on the CMU PIE database

	Absolute Improvement in the Correct Recognition Rate (%)	Decrease in the Error Rate (%)
DCT-0	9.2	78.0
PCA-all	30.4	92.1
PCA-3	18.4	87.6

Acknowledgement

This work is sponsored by the European Union under the integrated project CHIL, contract number 506909.

References

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

[2] R. Brunelli, T. Poggio, "Face Recognition: Features versus Templates", IEEE Trans. on PAMI, 13(10): 1042-1052, 1993.

[3] J.H.P.N. Belhumeur, D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection", IEEE Trans. on PAMI, 19(7):711–720, 1997.

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

[5] A. Pentland, B. Moghaddam, T. Starner and M. Turk, "View based and modular eigenspaces for face recognition", Proceedings of IEEE CVPR, pp. 84-91, 1994.

[6] H.K. Ekenel, R. Stiefelhagen, "Local Appearance based Face Recognition Using Discrete Cosine Transform", EUSIPCO 2005, Antalya, Turkey, 2005.
[7] R. Gottumukkal, V.K. Asari, "An improved face recognition technique based on modular PCA approach", Pattern Recognition Letters*,* 25(4), 2004.
[8] B. Heisele et al, "Face recognition: component-based versus global approaches", Computer Vision and Image Understanding, 91:6-21, 2003.
[9] T. Kim et al, "Component-based LDA Face Description for Image Retrieval and MPEG-7 Standardisation", Image and Vision Computing, 23(7):631-642, 2005.
[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.
[15] H.K. Ekenel, R. Stiefelhagen, "A Generic Face Representation Approach for Local Appearance based Face Verification", CVPR IEEE Workshop on FRGC Experiments, 2005.
[16] H.K. Ekenel, A. Pnevmatikakis, "Video-based Face Recognition Evaluation in the CHIL Project –Run 1", 7th Intl. Conf. on Automatic Face and Gesture Recognition (FG 2006), Southampton, UK, April 2006.
[17] A.M. Martinez, R. Benavente, "The AR Face Database", CVC Technical Report #24, 1998.
[18] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database", Proc. of Intl. Conf. On Automatic Face and Gesture Recognition, 2002.
[19] H.K. Ekenel, R. Stiefelhagen, "Analysis of Local Appearance-based Face Recognition on FRGC 2.0 Database", Face Recognition Grand Challenge Workshop (FRGC), Arlington, VA, USA, March 2006.