# Block Selection in the Local Appearance-based Face Recognition Scheme

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 we investigate the benefits of using a local appearance-based face recognition scheme against the problem of facial occlusion. We proposed two separate automatic block selection approaches to select the local image blocks that could be used for classification. Proposed approaches are tested against both upper and lower facial occlusions using the AR face database. Significant improvements are observed in the face recognition performance.*

## 1. Introduction

Face recognition has become one of the most addressed pattern recognition problems due to its importance as a natural biometric trait and due to its role in human computer interfaces.

Holistic approaches have been dominating the face recognition research since the beginning of 1990s [1-4]. On the other hand, recently, local appearance based face recognition approaches have attracted a growing interest [5-9]. In [5] salient local regions, such as the eye regions, are used to perform modular eigenfaces based face recognition. In [6], the face image is divided into rectangular smaller sub-images without considering any specific regions, and the eigenfaces approach is then performed on each of these sub-images. In [7], the local facial regions are located by a support vector machine (SVM) and the combined local features are classified again with SVM. In [8], the face image is partitioned into several local regions and each local region is represented by linear discriminant analysis (LDA). To combine the features extracted from each local region, another LDA is used.

In [9], a generic face representation approach is introduced as a baseline for local appearance based face recognition. Discrete cosine transform (DCT) is utilized for representing the local regions. 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 the 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 [10] and Yale [3] face databases. It is compared with the well known holistic approaches Eigenfaces [1], Fisherfaces [3], two face recognition architectures of independent component analysis (ICA) [4], and with the other local appearance based method that uses principal component analysis (PCA) [6]. The experimental results show that the proposed local appearance based approach performs significantly better than the holistic approaches. It also outperforms modular PCA approach [6] which indicates that DCT is a better choice than the PCA for local appearance-based face representation. Besides the performance improvement, the proposed approach has the advantages of using data independent basis and fast computation of the DCT features. Moreover, this approach is tested on FRGC version 1 data set for face verification [11], and a recent version of it on FRGC version 2 data set for face recognition [12], and it provided better and more stable results than the holistic baseline –eigenfaces. This representation scheme is also tested under video-based face recognition evaluations and again provided better results [13,14].

In this paper, following the studies [9,11,12,13,14], we investigate the effect of block selection to the performance of local appearance based face recognition scheme. One of the benefits of using the local appearance-based face recognition scheme lies in its robustness against local variations. In a holistic appearance-based approach a local variation due to expression, illumination or occlusion changes can modify the entire feature vector, however, in a local appearance-based scheme a local variation effects only the feature coefficients extracted from corresponding local regions. Furthermore, the local appearance-based scheme facilitates selection of the “important” local regions for face recognition. This way, both the performance and speed of face recognition can be increased by using only these “important” local regions. We proposed two block selection approaches that select the “important” local regions automatically for each test image. That is, for each test image, a different set of image blocks are used according to the result of block selection algorithm. This approach facilitates to handle the local variations that occur on different parts of the face image. For instance, it is expected from the block selection algorithm to choose the local regions located at the upper facial part if there is a variation at the lower part of the face, or vice versa.

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 proposed block selection approaches 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 used especially in compression standards due to its compact representation power. It's known that Karhunen-Loeve transform (KLT) is the optimal transform in terms of information packing, however, its data dependent nature makes it infeasible to implement in some practical tasks. Moreover, 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.

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

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

The obtained DCT basis functions for $N = 4$ are illustrated in Fig. 1 (each base is scaled separately for illustration purposes). As one can notice 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.

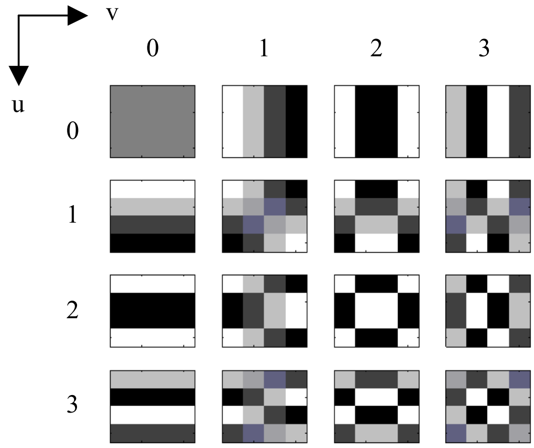


**Figure 1.** An illustration of DCT basis functions for N = 4

## 3. Face Recognition Using DCT

Local appearance based face representation using block-based discrete cosine transform 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 [5,7,8] for face representation. The local appearance based face representation can be performed as follows: A detected and normalized face image is divided into blocks of 8x8 pixels size. 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. It is also the block size in the JPEG compression standard. On each block DCT is performed. The obtained DCT coefficients are ordered using the zig-zag scanning (Fig. 2).

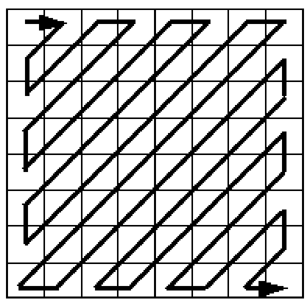

**Figure 2.** An illustration of zig-zag scan pattern

From the ordered coefficients, 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.

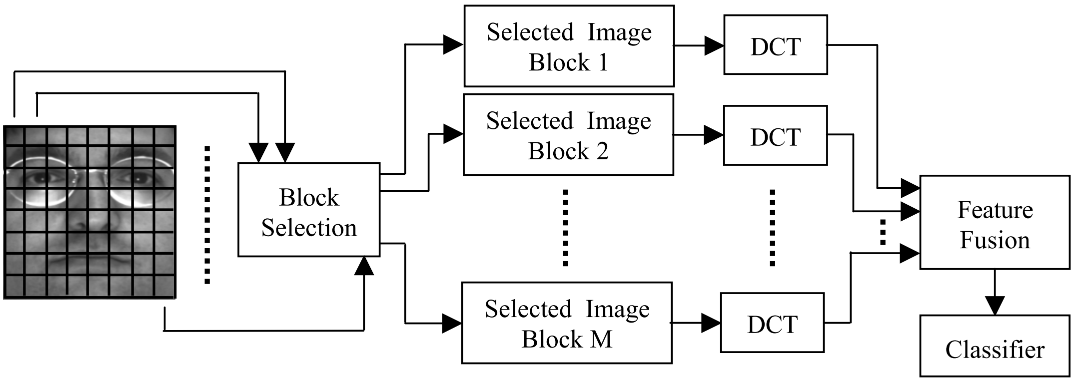


**Figure 3.** The diagram of block selection using image pixel values

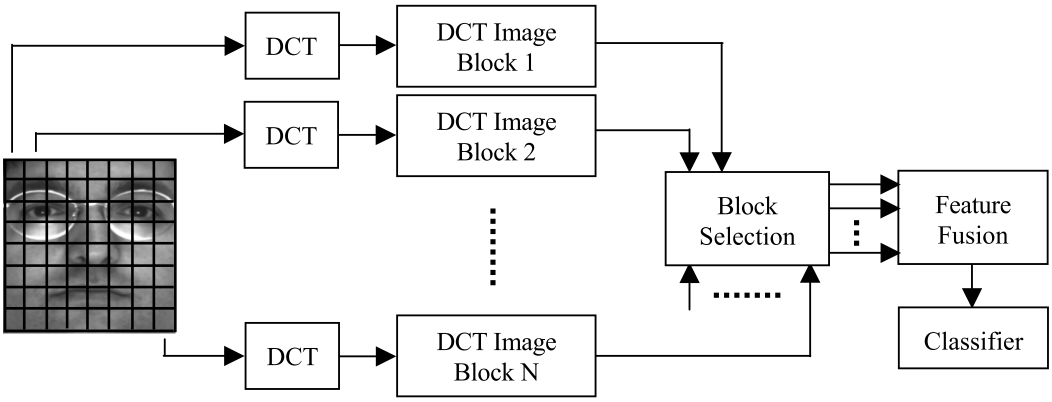


**Figure 4.** The diagram of block selection using DCT coefficients

## 4. Block Selection

Block selection is done by measuring the similarity between the pixel values or DCT coefficients of each block of the average training image and the pixel values or DCT coefficients of the corresponding blocks of the test image (Fig. 3,4). This facilitates adaptive block selection. In other words, for each new test face image, different blocks can be selected.

Block selection using image pixel values (Fig. 3) is performed as follows. The average pixel values of the i$^{th}$ block of the training images can be calculated as

$$\boldsymbol{b}_{tra,i} = \frac{1}{N}\sum_{l=1}^{N}\boldsymbol{b}_{tra,i,l} \ , \tag{3}$$

where $N$ is the number of images in the training set and $\boldsymbol{b}_{tra,i}$ represents the vector that contains the average pixel values of the i$^{th}$ block. This $K$ dimensional vector's mean value can be calculated as

$$m_{tra,i} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{b}_{tra,i,k} \ . \tag{4}$$

By subtracting the vector's mean value from its pixel values, the $\overline{\boldsymbol{b}}_{tra,i}$ is obtained

$$\overline{\boldsymbol{b}}_{tra,i} = \boldsymbol{b}_{tra,i} - m_{tra,i} \ . \tag{5}$$

Similarly, the mean value of the i$^{th}$ block in the test image can be calculated as

$$m_{test,i} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{b}_{test,i,k} \ . \tag{6}$$

Again, by subtracting the vector's mean value from its pixel values, the $\overline{\boldsymbol{b}}_{test,i}$ is obtained

$$\overline{\boldsymbol{b}}_{test,i} = \boldsymbol{b}_{test,i} - m_{test,i} \ . \tag{7}$$

Finally, the similarity between the i$^{th}$ block of the average training image and the corresponding block of the test image is calculated as

$$s_i = \bar{\boldsymbol{b}}_{tra,i} * \bar{\boldsymbol{b}}_{test,i} / \|\bar{\boldsymbol{b}}_{tra,i}\| * \|\bar{\boldsymbol{b}}_{test,i}\| \,. \tag{8}$$

These calculated similarity scores are ordered and the blocks that have higher scores are used for face recognition.

Block selection using DCT coefficients (Fig. 4) can be performed by following the same equations. In these equations, instead of using the image pixel values, the extracted M-dimensional local DCT feature vectors should be used. In this way, block selection procedure can be speeded up significantly.

## 5. Experiments

Two separate experiments are conducted to assess the contribution of block selection to the performance of the local appearance based face recognition algorithm. In the first experiment, the face recognition system is trained with face images that have no occlusion and tested against occluded faces from the same recording session. In the second test, again the system is trained with the face images that have no occlusion, but this time tested against occluded faces from a different recording session. The system is tested separately against two different kinds of occlusions: Person wearing sun-glasses and person wearing a scarf. The local appearance based approach is also compared with the holistic baseline –eigenfaces [1]. The AR face database [15] is used in the experiments.

The nearest neighborhood classifier is used in the study. The $L_1$ norm is used as the distance metric since it is observed that, it provides better results in both of the local and holistic approaches compared to the $L_2$ norm and cosine angle. The $L_1$ distance between $M$ dimensional feature vectors $f_{training}$ and $f_{test}$ can be calculated as

$$L1:\ d = \sum_{m=1}^{M} \left| f_{training,m} - f_{test,m} \right| . \tag{9}$$

### 5.1. Intra-session Experiments

The face database used in the intra-session experiments consists of 990 face images of 110 individuals that are taken from the first session of the AR face database [15]. Each individual in the derived face database has nine images. These images are annotated as “1: neutral expression”, “5: left light on”, “6: right light on”, “8: wearing sun glasses”, “9: wearing sun glasses and left light on”, “10: wearing sun glasses and right light on”, “11: wearing scarf”, “12: wearing scarf and left light on”, “13: wearing scarf and right light on”. From these nine images, the ones with annotations “1, 5, 6” are used for training and the remaining ones with annotations “8, 9, 10” are used for testing against upper face occlusion and the ones with annotations “11, 12, 13” are used for testing against lower face occlusion. 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

As stated in Section 3, the aligned face images are divided into 8x8 pixels blocks resulting in 64 non-overlapping local image regions as shown in Fig. 6. According to the used block selection approach, either block selection is performed first and then DCT is applied on the selected blocks (Fig. 3) or DCT is applied first and then block selection is done (Fig. 4). For classification, 5-dimensional local feature vectors are extracted from each selected block. For the baseline eigenfaces system 320 dimensional feature vectors are used for classification which is also the dimension of feature vectors when all the blocks are used for classification in a local appearance-based face recognition scheme.

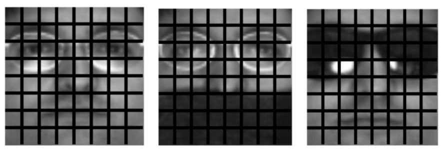

**Figure 6.** Sample partitioned face images

The correct classification rates of the block selection approaches against lower face occlusion problem are shown in Fig. 7. DCT –BS 1 denotes block selection using image pixel values and DCT –BS 2 implies block selection using DCT feature vectors. For the local appearance-based scheme, this figure shows the correct recognition rates for each used number of blocks. For eigenfaces approach, it shows the correct recognition rate using 320 dimensional feature vectors. It is apparent that both of the block selection approaches and the local appearance-based scheme without block selection outperform the eigenfaces significantly. The correct recognition rate obtained by eigenfaces is very low, however, this result is very close to the one obtained in [16], showing the difficulty of face recognition against occlusion.

As it can be observed from Fig. 7, block selection using image pixel values perform slightly better then block selection using DCT coefficients, nevertheless the performance difference is not significant. Therefore if the processing time is a critical issue, then instead of using 64-dimensional image pixels for similarity measurement, one can use 5-dimensional DCT feature vectors without sacrificing much from the face recognition performance.

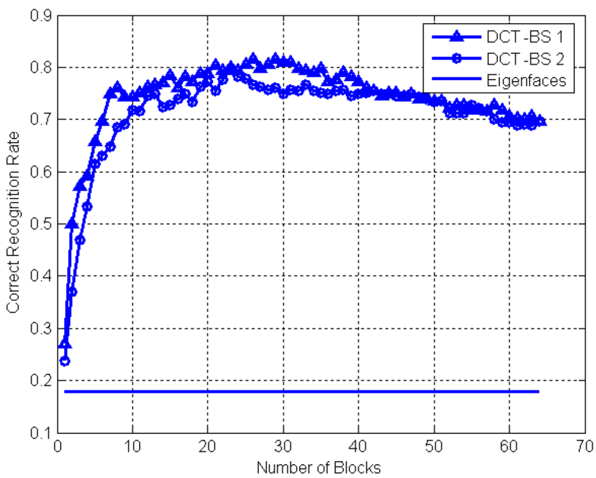


**Figure 7.** Correct recognition rate versus number of used blocks on the intra-session experiment with lower face occlusion

The performance increases in both of the block selection approaches at the beginning as the used number of blocks increases. It saturates when the number of used blocks is around 10. After this point it increases slightly till around half of the blocks are used. The performance decreases by further increase in the number of used blocks. For instance, the correct recognition rate when one uses only 10 blocks with the first block selection approach is 74.2%, it is 79.7% when one uses 32 blocks, and 69.7% when all the blocks are used. This shows that block selection contributes to the face recognition performance significantly. Moreover, it also speeds up the classification process by halving the feature vector's dimensionality. Although, local appearance-based face recognition approach with a block selection scheme requires an additional step of determining the "important" blocks, when the number of training samples that a test image should be compared against increases, this extra calculation remains insignificant compared to the total feature comparison during classification. Let's say we have $N$ training images in our database that a test image should be compared against, which results $DxN$ feature point comparisons, where $D$ is the dimension of the feature vector. If we use DCT coefficients for block selection, it causes only an additional comparison of the test image with the average training image, which results in total $D*(N/2+1)$ feature point comparisons assuming that we only use the half of the number of total blocks for classification. Since, in general, the number of training samples we have is much more than two, $N >> 2$, the amount of processing time required for face recognition using local appearance-based approach with block selection is almost half of the required processing time for local appearance-based face recognition without block selection.

In Fig. 8, average importance order of the blocks that is obtained against lower face occlusion is shown. Smaller numbers imply more importance. This order is obtained by sorting the averaged block similarity measures calculated for each test face image. It can be seen that, as one can expect, the blocks located in the upper half of the face image have more importance.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 8 | 20 | 21 | 31 | 34 | 33 | 24 | 1 |
| 39 | 32 | 19 | 7 | 6 | 17 | 30 | 40 |
| 9 | 5 | 13 | 18 | 14 | 12 | 2 | 3 |
| 23 | 26 | 29 | 4 | 11 | 22 | 10 | 28 |
| 38 | 56 | 35 | 15 | 16 | 44 | 58 | 37 |
| 55 | 25 | 42 | 57 | 54 | 36 | 27 | 53 |
| 48 | 63 | 49 | 51 | 52 | 45 | 64 | 47 |
| 60 | 43 | 46 | 61 | 59 | 50 | 41 | 62 |

**Figure 8.** The average importance order of the blocks obtained on the intra-session experiment with lower face occlusion

Fig. 9 shows the performance of the block selection approaches against upper face occlusion problem. Again, the block selection approaches perform much better than the baseline, eigenfaces approach. However, this time, the correct recognition rates obtained by local appearance-based scheme are not as high as the ones obtained against lower face occlusion. This indicates that, upper face regions contain more discriminative information than that of the lower face regions.

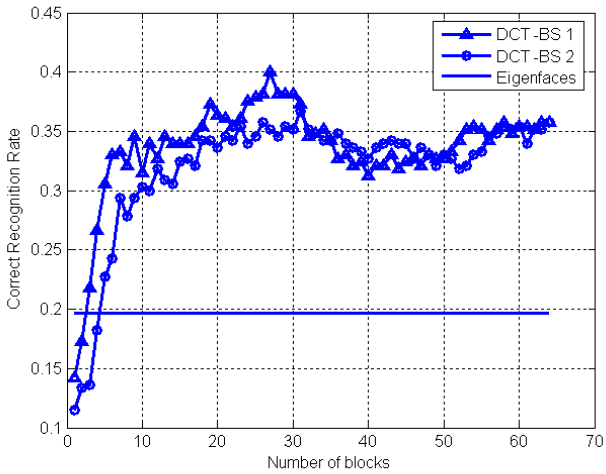


**Figure 9.** Correct recognition rate versus number of used blocks on the intra-session experiment with upper face occlusion

The average importance order of the blocks obtained against upper face occlusion is given in Fig. 10. As it can be observed, the blocks that correspond to the sun-glasses regions have the least importance. This validates our adaptive block selection scheme, that is, the location of the selected blocks changes as the location of variation changes. This is a very important property. If a system can not adapt the weights of the local regions, it can not handle the variations that may occur on the different parts of the face. For example in [16], the tested commercial face recognition software performed very well against lower face occlusion, however, its face recognition performance was less than half of the baseline eigenfaces' performance when it is tested against upper face occlusion. This implies that within the commercial software, the upper face regions were assigned with higher fixed weights than the lower face regions which caused poor results for the upper face occlusion case.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 6 | 20 | 23 | 29 | 28 | 30 | 22 | 3 |
| 56 | 50 | 7 | 16 | 25 | 10 | 44 | 51 |
| 57 | 58 | 59 | 62 | 55 | 60 | 61 | 41 |
| 63 | 43 | 54 | 11 | 4 | 64 | 46 | 48 |
| 38 | 8 | 17 | 26 | 33 | 39 | 1 | 37 |
| 2 | 19 | 24 | 45 | 35 | 15 | 27 | 5 |
| 14 | 42 | 36 | 34 | 31 | 32 | 49 | 13 |
| 53 | 18 | 12 | 47 | 40 | 9 | 21 | 52 |

**Figure 10.** Average importance order of the blocks obtained on the intra-session experiment with upper face occlusion

## 5.2. Inter-session Experiments

In the inter-session experiments, the same amount of data is used, that is, the used face database consists of 990 face images of 110 individuals that are taken from the AR face database [15]. Each individual in the derived face database has nine images. The same training data is used, as the one used in intra-session experiments which are selected from the first recording session of the AR face database. On the other hand, the test images are selected from a different session which is separated by two weeks from the first recording session. They have the same annotations with the ones in intra-session experiments. The face images are aligned using the eye center locations and scaled to 64x64 pixels resolution.

Fig. 11 plots the correct recognition rates of the block selection approaches against lower face occlusion for varying number of used blocks. The observed outcomes are similar to the ones obtained in intra-session experiments. The main difference, one can notice is the relatively worse performance values. This is expected, since in this experiment the occlusion problem is coupled with the time gap between training and testing data, hence causing a more difficult problem.

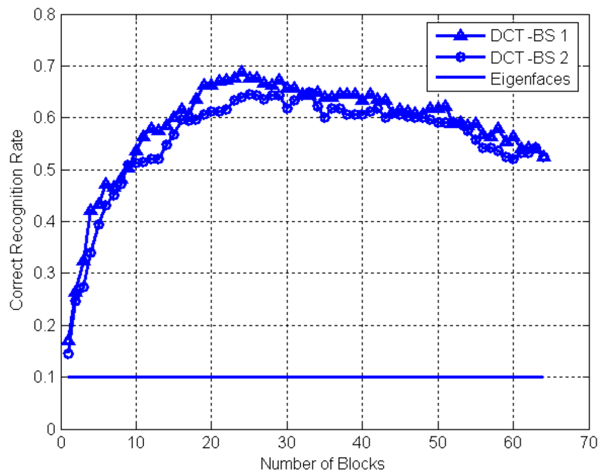


**Figure 11.** Correct recognition rate versus number of used blocks on the inter-session experiment with lower face occlusion

The average importance order obtained in this experiment is depicted in Fig. 12. Although there is a change in the order, the blocks located in the upper half of the face image still have more importance. This verifies our block selection scheme.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4 | 20 | 19 | 24 | 29 | 32 | 25 | 3 |
| 37 | 31 | 18 | 6 | 5 | 12 | 26 | 40 |
| 10 | 8 | 11 | 21 | 15 | 14 | 2 | 7 |
| 22 | 27 | 30 | 1 | 9 | 23 | 16 | 28 |
| 48 | 50 | 41 | 17 | 13 | 53 | 51 | 42 |
| 57 | 33 | 44 | 58 | 59 | 36 | 34 | 52 |
| 39 | 64 | 47 | 54 | 56 | 49 | 63 | 46 |
| 55 | 35 | 38 | 61 | 60 | 43 | 45 | 62 |

**Figure 12.** Average importance order of the blocks obtained on the inter-session experiment with lower face occlusion

Correct recognition rates and importance order of the blocks against upper facial occlusion is shown in Figs. 13 and 14, respectively. Again, the results of this experiment confirm the findings of the intra-session experiments.

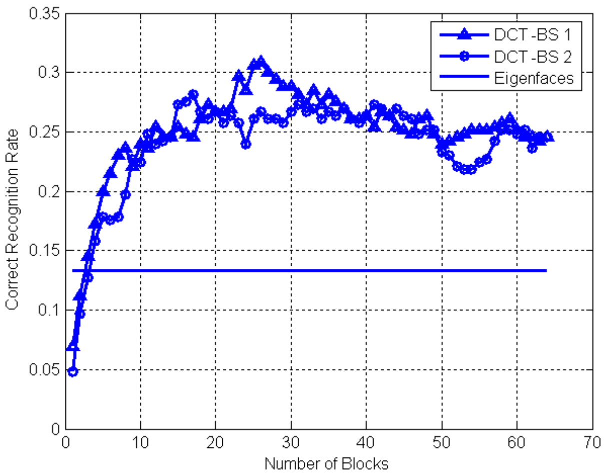


**Figure 13.** Correct recognition rate versus number of used blocks on the inter-session experiment with upper face occlusion

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3 | 25 | 21 | 24 | 28 | 32 | 17 | 1 |
| 58 | 51 | 5 | 14 | 23 | 9 | 38 | 50 |
| 54 | 60 | 57 | 62 | 52 | 61 | 59 | 41 |
| 64 | 44 | 56 | 10 | 6 | 63 | 36 | 49 |
| 45 | 7 | 18 | 22 | 34 | 43 | 2 | 39 |
| 4 | 26 | 29 | 42 | 37 | 16 | 27 | 8 |
| 12 | 46 | 31 | 35 | 30 | 33 | 48 | 11 |
| 55 | 19 | 15 | 47 | 40 | 13 | 20 | 53 |

**Figure 14.** The average importance order of the blocks obtained on the inter-session experiment with upper face occlusion

## 6. Conclusions

In this paper we investigate the benefits of using adaptive block selection in a local appearance-based face recognition scheme. We proposed two block selection approaches and perform extensive experiments against lower and upper face occlusion. We observed that even without block selection, local appearance-based face recognition outperforms the holistic baseline significantly against the occlusion problem. With block selection, only by using half of the total number of blocks, around 10 to 15% absolute increase is obtained in the correct recognition rate against lower face occlusion, and around 5% absolute increase is obtained in the correct recognition rate against upper face occlusion. Moreover, the reduction in feature dimensionality speeds up the face recognition system.

As a future work, we plan to develop a probabilistic weighting scheme to weight each block's contribution to the face classification.

## Acknowledgements


We would like to thank Ralph Gross for providing us the eye center labels of the AR face database. We also would like to thank Dr. Jie Yang for his contributions.

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

[7] B. Heisele et al, "Face recognition: component-based versus global approaches", Computer Vision and Image Understanding, 91:6-21, 2003.

[8] 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.

[9] H.K. Ekenel, R. Stiefelhagen, "Local Appearance based Face Recognition Using Discrete Cosine Transform", EUSIPCO 2005, Antalya, Turkey, 2005.

[10] 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.

[11] H.K. Ekenel, R. Stiefelhagen, "A Generic Face Representation Approach for Local Appearance based Face Verification", CVPR IEEE Workshop on FRGC Experiments, 2005.

[12] 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.

[13] 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.

[14] H.K. Ekenel, Q. Jin, "ISL Person Identification Systems in the CLEAR Evaluations", CLEAR Evaluation Workshop, Southampton, UK, April 2006.

[15] A.M. Martinez, R. Benavente, "The AR Face Database", CVC Technical Report #24, 1998.

[16] R. Gross et al, "Que vadis face recognition*", 3rd* Workshop on Emprical Evaluation Methods in Computer Vision, 2001.