

# Why is Facial Occlusion a Challenging Problem?

Hazım Kemal Ekenel and Rainer Stiefelhagen

Computer Science Department, Universität Karlsruhe (TH)  
Am Fasanengarten 5, Karlsruhe 76131, Germany  
{[ekenel](mailto:ekenel@ira.uka.de), [stiefel](mailto:stiefel@ira.uka.de)}@ira.uka.de  
<http://isl.ira.uka.de/cvhci>

**Abstract.** This paper investigates the main reason for the obtained low performance when the face recognition algorithms are tested on partially occluded face images. It has been observed that in the case of upper face occlusion, missing discriminative information due to occlusion only accounts for a very small part of the performance drop. The main factor is found to be the registration errors due to erroneous facial feature localization. It has been shown that by solving the misalignment problem, very high correct recognition rates can be achieved with a generic local appearance-based face recognition algorithm. In the case of a lower face occlusion, only a slight decrease in the performance is observed, when a local appearance-based face representation approach is used. This indicates the importance of local processing when dealing with partial face occlusion. Moreover, improved alignment increases the correct recognition rate also in the experiments against the lower face occlusion, which shows that face registration plays a key role on face recognition performance.

## 1 Introduction

Partial face occlusion is one of the most challenging problems in face recognition. A face recognition system can confront occluded faces in real world applications very often due to use of accessories, such as scarf or sunglasses, hands on the face, the objects that persons carry, and external sources that partially occlude the camera view. Therefore, the face recognition system has to be robust to occlusion in order to guarantee reliable real-world operation. Sample occluded face images from a real world face recognition application are shown in Figure 1.

Several studies have been conducted in order to address this problem [1–6]. In [1], face images are analyzed locally in order to handle partial face occlusion. The face image is first divided into  $k$  local regions and for each region an eigenspace is constructed. If a region is occluded, it is automatically detected. Moreover, weighting of the local regions are also proposed in order to provide robustness against expression variations. A similar approach is presented in [2], where a self-organizing map (SOM) is used to model the subspace instead of Gaussians or mixtures of Gaussians as in [1]. A face is represented by the face attributed relational graph (ARG) structure in [3]. This representation contains a set of



**Fig. 1.** Sample occluded face images.

nodes and binary relations between these nodes. In testing, first the correspondences between the ARG representations of the training and testing samples are established. According to the distance between these representations, the classification is performed. In [4], robustness against occlusion is provided by combining the subspace methods that aim at best reconstruction, such as principal component analysis, with the subspace methods that aim at discrimination, such as linear discriminant analysis. A sparse signal representation is used to analyze partially occluded face images in [5]. Another representation based approach is proposed in [6]. Different from the studies [1–5] in which occluded images are only included in the testing set, in [6] they are included both in the training and testing sets. The occlusion problem is handled as a reconstruction problem and the classification is done according to the obtained reconstruction error on a test image.

All these studies have focused on finding a way to handle occlusion, especially upper face occlusion. However, until now, no study has been conducted on finding the main reason of low performance caused by the occlusion. Mainly, degraded performance is attributed to the fact that occlusion results in loss of discriminative information, especially in the case of upper face occlusion the eye region, which is known to be very discriminative, can not be utilized. In this study, we thoroughly analyzed the occlusion problem via various experiments and found that the main reason of low performance is not due to missing discriminative information, but the registration errors.

## 2 Local appearance-based face recognition

In this study, among the generic face recognition algorithms, local appearance-based face recognition algorithm [7, 8] is chosen for face recognition. The reason to opt for this algorithm over other generic face recognition algorithms, such as eigenfaces [9] or Fisherfaces [10], is that it utilizes local facial regions for feature extraction, which is also a common characteristic of many face recognition algorithms that have been developed specifically to handle facial occlusion.

In the local appearance-based face recognition (LAFR) approach, a detected and registered face image is divided into blocks of  $8 \times 8$  pixels size. Afterwards, the discrete cosine transform (DCT) is performed on each  $8 \times 8$  pixels block. The obtained DCT coefficients are ordered using zig-zag scanning. From the ordered

coefficients, according to a feature selection strategy,  $M$  of them are selected and normalized, resulting in an  $M$ -dimensional local feature vector.

Regarding feature normalization, there are two aspects. The first aspect is the value range of the DCT coefficients. The first coefficients have higher magnitudes than the later ones. Therefore, they contribute more to the calculated distance in a nearest neighbor classification scheme, hence having more importance in the classification. However, it is known that being able to represent more energy does not necessarily imply to have more discriminative power [11]. The second aspect is the total magnitude of each block’s DCT coefficients. Since DCT is an orthonormal transformation and conserves all the energy of the processed input blocks, blocks with different brightness levels lead to DCT coefficients with different value levels. Because of this reason, the blocks with brighter content have more impact on the classification results. In order to balance the coefficients’ contributions to the classification, the coefficients in each feature vector are first divided by their standard deviations that are learned from the training set, and then to balance the blocks’ contributions to the classification, the local feature vector is normalized to unit norm [8].

Finally, the DCT coefficients extracted from each block are concatenated to construct the overall feature vector. The classification is done using a nearest neighbor classifier with L1 norm as the distance metric.

### 3 How much performance loss does occlusion cause?

Besides the face recognition algorithms that have been specifically developed to handle occlusion, face recognition algorithms attain very low performance when they are evaluated on occluded face images. For example, in [12], in the experiments on the AR face database [13], eigenfaces algorithm [9] obtained 48%, Fisherfaces [10] 45%, and FaceIt [14] 10% correct recognition rate when they were tested on the face images that contain upper face occlusion caused by sunglasses, whereas they attained 27%, 44%, and 81% respectively when they were tested on the face images that contain lower face occlusion caused by scarf.

In order to assess the drop in face recognition algorithm’s performance due to partial face occlusion, we artificially occlude the eye region of the aligned faces and run the local appearance-based face recognition algorithm on the occluded faces, as well as on the face images without occlusion. The occlusion is imitated by painting the blocks in the second and third rows black. A sample aligned face image and the corresponding occluded face image are depicted in Figure 2.



**Fig. 2.** Sample aligned face image and corresponding occluded face image.

The experiments have been conducted on the *Face Recognition Grand Challenge* (FRGC) database [15]. For each case, we run two different experiments. One, *FRGC1*, with face images collected under controlled conditions and one, *FRGC4*, with face images collected under uncontrolled conditions. For controlled conditions, we selected 120 individuals from the FRGC experiment 1 data set, who have at least ten images both in spring 2003 and fall 2004 recordings. We used the images from spring 2003 for training and the ones from fall 2004 for testing. For uncontrolled conditions, we had the same experimental setup, but this time the images were selected from the FRGC experiment 4 data set. Table 1 shows the experimental setup. Sample images from the data sets can be seen in Figure 3.

Label of the experiment	Gallery	Probe
FRGC1	Face images collected under controlled conditions during fall 2003	Face images collected under controlled conditions during spring 2004
FRGC4	Face images collected under uncontrolled conditions during fall 2003	Face images collected under uncontrolled conditions during spring 2004

**Table 1.** Experiments on the FRGC database.



**Fig. 3.** Sample face images from the FRGC database. The first two images from the left are samples of face images collected under controlled conditions. The remaining two images are samples of face images collected under uncontrolled conditions.

The classification results with this experimental setup is given in Table 2. The results obtained on the test face images without occlusion are also presented in the table for comparison purposes. It is apparent that missing eye region information causes a decrease in the correct classification rate. Especially, if the experiment is more difficult, as in the FRGC4 experiment, where the training and testing data is collected under uncontrolled conditions, the decrease is more prominent. However, compared to the results obtained on the AR face database [12], the results are still high and the performance drop is not that dramatic. This finding implies that, it is not only the missing eye region information that causes degradation in performance, but also there are some other factors which will be investigated in the next section.

	FRGC1	FRGC4
Without occlusion	97.9%	90.8%
With occlusion	95.9%	83.8%

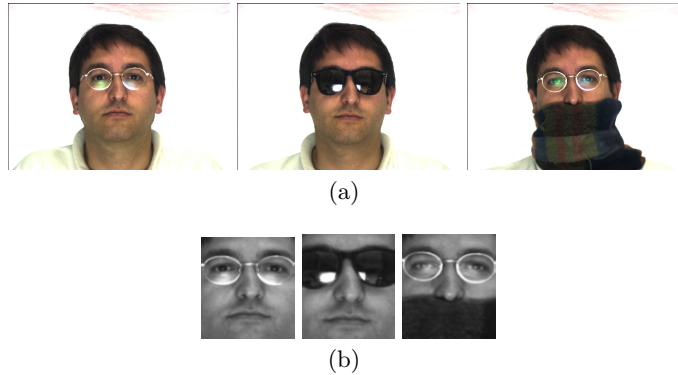
**Table 2.** Correct recognition rates obtained on the FRGC database. The results in the first row is obtained using test face images that have no occlusion. The ones in the second row is attained using test face images that contain upper face occlusion as depicted in Figure 2.

## 4 What is the main problem?

In this section, we conduct experiments on the occluded face images, compared the results with the ones from the previous section and try to find out the main problem. We used the AR face database [13] for the experiments. From the database, one image per subject is used from the first session for training. This image is annotated as “1: neutral expression”. For testing we used five images per subject. Face images from the second session, which are annotated as “14: neutral expression”, are used to assess the algorithm’s performance when there exists no occlusion. For testing on occluded face images, two images per subject are used from each session, which are annotated as “8/21: wearing sunglasses” and “11/24: wearing scarf”, where the first number corresponds to the label in the first recording session and the second one corresponds to the label in the second recording session. From these two images, the ones with annotations “8/21: wearing sunglasses” are used for testing against upper face occlusion and the ones with annotations “11/24: wearing scarf” are used for testing against lower face occlusion. In the data set, there are 110 subjects who have all these samples in both of the sessions. Overall, five separate experiments are conducted on this data set. Setups of the experiments are presented in Table 3. Sample input images and corresponding registered images are shown in Figure 4. Face images are aligned with respect to manually labeled eye center locations.

Label of the experiment	Gallery	Probe
AR1scarf	Face images without occlusion from session 1	Face images with scarf from session 1
AR1sun	Face images without occlusion from session 1	Face images with sunglasses from session 1
ARneutral	Face images without occlusion from session 1	Face images without occlusion from session 2
AR2scarf	Face images without occlusion from session 1	Face images with scarf from session 2
AR2sun	Face images without occlusion from session 1	Face images with sunglasses from session 2

**Table 3.** Experiments on the AR face database.



**Fig. 4.** Sample images from the AR face database. (a) Sample input images. (b) Corresponding registered images.

The results of these experiments are listed in Table 4. As can be observed, upper face occlusion causes a significant drop in performance. The correct recognition rate decreases from 92.7% to 37.3%. However, the performance decrease in the case of lower face occlusion is small. Compared to the results reported in [12], the performance obtained in the case of upper face occlusion is similar, on the other hand the results obtained on lower face occlusion is significantly higher. The performance increase in the case of lower face occlusion can be explained by the local representation approach, in which a change in a local region affects only the features that are extracted from the corresponding block while the features that are extracted from the other blocks remain unaffected, whereas in a holistic appearance-based face recognition approach, such as eigenfaces [9] or Fisherfaces [10], it can affect the entire feature representation. Nevertheless, this scheme does not help to improve the results obtained on the face images that contain upper face occlusion. Compared to the results in the previous section, the performance drop due to upper face occlusion on the AR face database is significantly higher. This implies that, there is another underlying problem, in addition to missing eye region information.

	Correct recognition rate
AR1scarf	91.8%
AR1sun	38.2%
ARneutral	92.7%
AR2scarf	83.6%
AR2sun	37.3%

**Table 4.** Correct recognition rates obtained on the AR face database.

Recently, some studies have been conducted on the robustness of face recognition algorithms against registration errors [16, 17]. These studies have shown that face recognition algorithms' performance rely on face alignment accuracy. In [16], the robustness of eigenfaces algorithm [9] is analyzed against horizontal and vertical translations, as well as scale and rotation, which might occur due to alignment using erroneously detected facial feature locations. It is found that, the eigenfaces approach can tolerate variations only upto 5% of the face image size. Similarly in [17], in addition to the eigenfaces approach, Fisherfaces [10], elastic bunch graph matching [18], Laplacianfaces [19], minimum average correlation energy (MACE) [20], and pseudo two-dimensional hidden Markov models [21] have been analyzed against registration errors. All these algorithms are observed to be sensitive to misalignment. To address this problem, in [22] a face alignment algorithm is proposed. In the proposed approach, face registration is done by minimizing the closest distance at the classification step. In this approach, once the face is located, positions of the facial features are roughly estimated. Search for the precise facial feature positions is conducted around the estimated positions. Various candidate facial feature positions are used to provide several aligned test face images, while comparing a test face image with an already aligned training face image. The facial feature positions, which lead to the aligned test face image that has the minimum distance to the training image, are selected as the facial feature locations. Thus for each training sample, separate eye center positions are determined for the test face image. In this way, inconsistencies across manual eye center labels of the training images are also handled.

Considering the significant performance drop in the case of upper face occlusion compared to the case of lower face occlusion and compared to the case of artificial occlusion, we hypothesize that the possible reason for this outcome is misalignment. Since the faces are occluded, even manual eye center labels can not be precise resulting misalignment. On the other hand, when we occlude already aligned face images we do not have this problem and the only factor that causes performance drop is the missing eye region information. In order to validate this hypothesis, we used the alignment algorithm in [22] which has been shown to be insensitive to misalignment due to imprecise facial feature localization. The results of this experiment are given in Table 5. It can be seen that the performance increases significantly. Even though a generic face recognition algorithm is used, the results are even better than the state-of-the-art algorithms that have been specifically developed to handle occlusion [1–6]. It is also interesting to observe that, even the lower face benefit from better alignment, which implies that manual eye center labels may not be consistent, especially when the face images are not very high resolution. Another interesting point is that, slightly higher correct recognition rates are achieved on the upper face occlusion than the lower face occlusion. The reason can be the textured surface of the scarfs, which might affect the classification decision more than the black sunglasses.

	Correct recognition rate
AR1scarf	97.3%
AR1sun	98.2%
ARneutral	100%
AR2scarf	93.6%
AR2sun	95.5%

**Table 5.** Correct recognition rates obtained on the AR face database with the improved face alignment.

## 5 Conclusions

In this paper, we investigated the main reason for the significant degradation of face recognition algorithms’ performance in the case of partial face occlusion. We first artificially generate occluded face images by painting the eye region black and performed experiments both on the face images collected under controlled and uncontrolled conditions. We found that eye region contains valuable discriminative information, especially in the case of uncontrolled conditions. Although, in this experiment we have observed a performance drop due to occlusion, the results were not as low as the ones reported in the literature against upper face occlusion. Afterwards, we conducted experiments on the AR face database [13], both on the lower and upper face occlusion. We found that lower facial occlusion can be handled by local processing. However, against upper face occlusion we got very low performance. This was very intriguing, since the performance drop was not that severe on the artificially occluded face images. Comparing the experimental conditions between the artificially occluded face images and the occluded face images from the AR face database, there are two main differences. The first one is, in the artificial data the eye region is completely black, whereas in the natural one there are reflections. The second difference is, the manual eye center labels, which have been used for alignment is more precise in the artificially occluded face images, whereas due to sunglasses, manual eye center labels of the face images from the AR face database is not reliable. Recent studies on the impact of registration errors on face recognition performance [16, 17] lead us to hypothesize that the second difference has more dominant effect. By utilizing an alignment approach [22], which is insensitive to imprecise facial feature localization, we run the experiments again and obtained very high increase in the performance indicating that the main reason causing the performance drop is not the missing eye region information, but the misalignment due to erroneously localized facial feature points. When it is handled, very high correct recognition rates can be obtained even with a generic local appearance-based approach.

## Acknowledgement

This study is partially funded by the German Research Foundation (DFG) under Sonderforschungsbereich SFB 588 —Humanoid Robots— and by OSEO, French



State agency for innovation, as part of the Quaero Programme. The authors would like to thank Mika Fischer for his contributions to the study.

## References

1. Martinez, A.M.: Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(6) (2002) 748–763
2. Tan, X., Chen, S., Zhou, Z.H., Zhang, F.: Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft  $k$ -NN ensemble. *IEEE Transactions on Neural Networks* **16**(4) (2005) 875–886
3. Park, B.G., Lee, K.M., Lee, S.U.: Face recognition using face-ARG matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(12) (2005) 1982–1988
4. Fidler, S., Skočaj, D., Leonardis, A.: Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(3) (2006) 337–350
5. Wright, J., Ganesh, A., Yang, A., Ma, Y.: Robust face recognition via sparse representation. Technical report, University of Illinois, USA (2007)
6. Jia, H., Martinez, A.M.: Face recognition with occlusions in the training and testing sets. In: *Proceedings of IEEE Int'l. Conf. on Automatic Face and Gesture Recognition*. (2008)
7. Ekenel, H.K., Stiefelhagen, R.: Local appearance-based face recognition using discrete cosine transform. In: *Proc. of 13<sup>th</sup> European Signal Processing Conf. (EU-SIPCO 2005)*. (2005)
8. Ekenel, H.K., Stiefelhagen, R.: Analysis of local appearance-based face recognition: Effects of feature selection and feature normalization. In: *Conf. on Computer Vision and Pattern Recognition Workshop 2006*. (2006) 34
9. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* **3**(1) (1991) 71–86
10. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class-specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7) (1997) 711–720
11. Ekenel, H.K., Sankur, B.: Feature selection in the independent component subspace for face recognition. *Pattern Recognition Letters* **25**(12) (2004) 1377–1388
12. Gross, R., Shi, J., Cohn, J.F.: Que vadis face recognition? In: *Proceedings of Workshop on Empirical Evaluation Methods in Computer Vision*. (2001)
13. Martinez, A.M., Benavente, R.: The AR face database. Technical Report 24, CVC (1998)
14. Penev, P.S., Atick, J.: Local feature analysis: a general statistical theory for object representation. *Network: Computation in Neural Systems* **7** (1996) 477–500
15. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the Face Recognition Grand Challenge. In: *Proc. 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*. Volume 1., Los Alamitos, CA, USA (2005) 947–954
16. Lemieux, A., Parizeau, M.: Experiments on eigenfaces robustness. In: *Proc. Int'l. Conf. on Pattern Recognition*. Volume 1. (2002) 421–424
17. Rentzperis, E., Stergiou, A., Pnevmatikakis, A., Polymenakos, L.: Impact of face registration errors on recognition. In: *Artificial Intelligence Applications and Innovations (AIAI06)*, Springer (2006) 187–194

18. Wiskott, L., Fellous, J.M., Kruger, N., Malsburg, C.: Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7) (1997) 775–779
19. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.J.: Face recognition using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(3) (2005) 328–340
20. Xie, C., Kumar, B.V.K.V., Palanivel, S., Yegnanarayana, B.: A still-to-video face verification system using advanced correlation filters. In: *Proceedings of International Conference on Biometric Authentication*. (2004) 102–108
21. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*. (1994) 138–142
22. Ekenel, H.K.: *A Robust Face Recognition Algorithm for Real-World Applications*. PhD thesis, Universität Karlsruhe (TH), Germany (2009)