

Face Alignment by Minimizing the Closest Classification Distance

Hazım Kemal Ekenel and Rainer Stiefelhagen

Abstract—In this paper, we present a face registration approach, in which alignment is done by minimizing the closest distance at the classification step. This method eliminates the need of a feature localization step that exists in traditional face recognition systems and formulates alignment as an optimization process during classification. In other words, instead of performing a separate facial feature localization step and localizing facial features according to some type of feature matching score, in the proposed method, alignment is done by directly optimizing the classification score. Moreover, a feature detector can still be integrated to the system. In this case, the output of the feature detector is used as the initial point of the optimization process. Results of extensive experiments have shown that the proposed approach leads very high correct recognition rates, especially in the case of partial face occlusion, where it is not possible to precisely detect the facial feature locations. It has been also found that, in the case of using a facial feature detector, the approach can tolerate localization errors of up to 18% of the interocular distance.

I. INTRODUCTION

Face alignment is one of the most crucial processing steps in face recognition systems. In order to have a meaningful comparison between two face images, they need to be aligned precisely. In traditional face recognition systems, to achieve this, after face detection, facial feature points—generally eye centers or corners—are located [1] and the face images are aligned with respect to these points. However, facial feature localization is a difficult task and imprecisely localized feature points lead to misaligned faces.

Recently, two studies have been published on the robustness of face recognition algorithms against registration errors [2], [3]. These studies have shown that face recognition algorithms' performance relies heavily on face alignment accuracy. In [2], the robustness of the eigenfaces algorithm [4] is analyzed against horizontal and vertical translations, as well as scale and rotation, which might occur due to alignment using erroneously localized facial feature points. It is found that the eigenfaces approach can tolerate variations only up to 5% of the face image size. Similarly in [3], in addition to the eigenfaces approach, Fisherfaces [5], elastic bunch graph matching [6], Laplacianfaces [7], minimum average correlation energy (MACE) [8], and pseudo two-dimensional hidden Markov models [9] have been analyzed

against registration errors. All these algorithms are observed to be sensitive to misalignment.

There are only a few studies that have focused on building face recognition systems which are robust to misalignment [10], [11], [12]. In [10], from each training sample, 6615 additional images are generated by perturbing the facial feature locations. These images are projected onto the eigenspace and for each subject the resulting feature vectors are modeled with a Gaussian or a mixture of Gaussians. The identification is done by finding the closest model. Similarly, in [11], 81 additional samples are derived from each training sample by modifying the eye center locations. These samples are then used as input to the Fisherfaces algorithm [5]. Different from these two studies, in which additional samples are generated, in [12], misalignment parameters are learned by solving an optimization problem. Although these approaches have improved performance over the baseline, they still cannot fully handle misalignment. For example, in [10], on the AR face database [13], around 80% and 50% correct recognition rates are obtained against occlusion over randomly selected 50 subjects, for the within session and between session experiments, respectively, whereas in this study around 95% and 90% correct recognition rates are achieved for the same conditions over 110 subjects. In [11], even with two pixels translation, the performance drops by 30%, whereas in this study, the performance stays constant with respect to translations up to 18% of the interocular distance. Similarly, in [12], there is still a significant performance drop even in the case of small modifications, such as rotation of $\pm 5^\circ$, scaling with [0.95, 1.05], or ± 1 pixel shift, whereas in this study rotations up to $\pm 30^\circ$, scaling factors up to [0.64, 1.36], and shifts up to 18% of the interocular distance, are tolerated. In our proposed approach, once the face is located with an automatic face detector, 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, since, as already mentioned, for each training sample a separate eye localization is performed by minimizing the classification distance. This is different from traditional face recognition approaches where only one eye center estimate is

This work was 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 Quæro Programme. The authors would like to thank Mika Fischer and Hua Gao for their contributions to the study.

The authors are with the Faculty of Computer Science, University of Karlsruhe (TH), 76131 Karlsruhe, Germany {ekenel, stiefel}@ira.uka.de

used to match the test image against all the training samples ignoring the possibility of having inconsistencies among the manual eye center labels of the training samples. Although the proposed system is free from feature localization, it is still possible to integrate a feature detector to it. In this case, the output of the feature detector is used as the initial point of the optimization process. Experimental results have shown once more that registration plays a very crucial role in face recognition systems. Using a registration approach that directly aims at minimizing the closest classification distance provides insensitivity to erroneous facial feature localization. In addition, it has been found that the main problem with the upper facial occlusion caused by sunglasses is not the missing eye region information, but the mis-localization of the eye-related feature points which have been widely used for face registration. Furthermore, in our experiments, the obtained results with the proposed alignment approach even outperform the ones obtained using the manual labels.

II. FACE ALIGNMENT

Since all humans have the same facial feature configuration, once the face is located it is easy to roughly estimate the locations of the facial features. In order to show this, we plot in Fig. 1 the relative eye center positions with respect to the center of the bounding boxes of faces. For this analysis, we use the training samples of the AR [13] and the FRGC [14] databases. The faces are detected with a generic face detector [15]. As can be observed from Fig. 1, despite using different databases and a generic automatic face detector, the normalized relative eye center positions are densely located. The median values of these eye center positions are calculated both for the left and right eye to produce an eye center hypothesis with respect to the automatically located face rectangle. In order to contain all the deviations from these calculated values due to variations in feature positions across different identities and variations in the bounding boxes generated by the automatic face detector, a window size of 11×11 pixels is determined around the eye centers. It can be seen in Fig. 1 that a window size of 9×9 would suffice to cover all the points. Furthermore, with a fine tuned face detector and improved face segmentation accuracy, this region can become even smaller. However, for the sake of having a generic approach without relying on the accuracy of the face detector and in order to tolerate in-plane and out-of plane rotations up to some extent, a larger window size is selected. These regions are used as search regions to determine the best matching eye centers between the test image and training samples. In order to save processing time during testing, in the implementation, the generation of aligned face images using the eye center coordinates within the determined region is performed offline, on the training side¹. The algorithm can be summarized as follows:

Training:

¹Although to save processing time we generated aligned face images on the training side, with a GPU programming implementation it is also possible to perform this processing real-time on the test side. For example, generating 705 aligned samples from a single test image takes only 24.8ms.

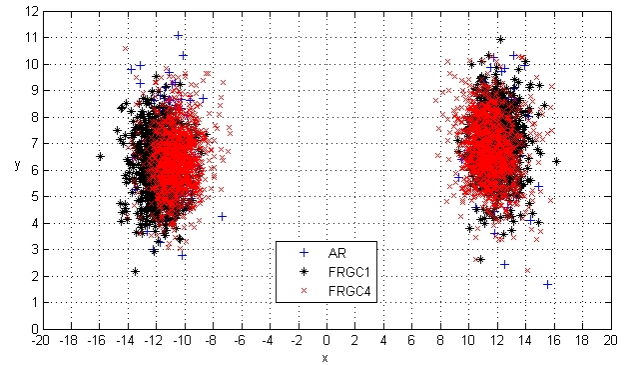


Fig. 1. Distribution of the eye centers with respect to the center of the face bounding box.

- (i) Have all the eye center location combinations between the left and right eyes within the 11×11 pixels window around the manual eye center labels,
- (ii) Generate aligned face images according to these eye center positions,
- (iii) Extract a feature vector from each aligned face image.

Testing:

- (i) Do face localization and estimate the eye center positions by adding $(-11.5, 6.5)$ for the left eye center and $(11.5, 6.5)$ for the right eye center to the center of the scaled face bounding box,
- (ii) Align test face image with respect to the estimated eye center position,
- (iii) Compare the aligned test face image with all the aligned face images generated from a training sample,
- (iv) Find the aligned face image from the training sample that provides the minimum classification distance,
- (v) Perform steps (iii) and (iv) for each training sample,
- (vi) Find the training sample that provides minimum classification distance,
- (vii) Assign the identity of the best matching training sample to the test image.

One can notice that having a window size of $W \times W$ pixels around the eye centers causes W^4 eye center position combinations. However, the amount of comparisons can be significantly decreased by utilizing a hierarchical search scheme. Instead of having all the position combinations within ± 5 pixels, first the combinations at ± 2 pixels locations can be searched. This way, the number of eye center position combinations at the first search step becomes 625. After determining the combination that provides the minimum classification distance at the first step, at the second step, the search is done ± 1 pixel around the determined eye center positions from the first search step. Thus, at the second step W becomes 3, providing 81 combinations. Since the classification is already done with the face image aligned using the determined eye center positions from the first search step, only 80 additional comparisons are needed to be done at the second step, making overall a total of 705 comparisons per training sample. This search pattern is

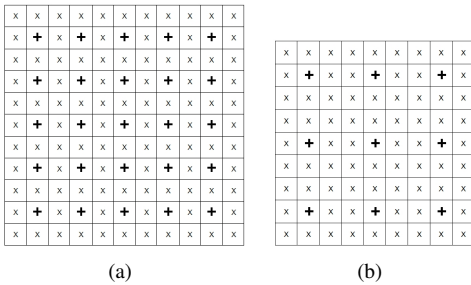


Fig. 2. Search pattern for (a) $W = 11$, (b) $W = 9$.

depicted in Fig. 2(a). The “+” shows the search locations at the first step, whereas “x” shows the search locations at the second step. The computational load can be further decreased by having a window size of $W = 9$, and performing the search first at the combinations of ± 3 pixels and then at the combinations of ± 1 pixel around the determined eye center locations from the first step, which makes in total 161 comparisons per training sample. The search pattern for $W = 9$ is shown in Fig. 2(b).

It should be noted that, although on one hand the amount of computation increases due to higher number of feature vector comparisons, on the other hand, due to omitting a separate feature detection step, some amount of computation is saved. Moreover, as will be presented in the next section, feature comparison consists of only a subtraction operation which can be performed very fast.

III. FACE RECOGNITION

In this study, among the generic face recognition algorithms, local appearance-based face recognition using the discrete cosine transform (DCT) [16] is chosen for face recognition. There are many advantages of this algorithm over other generic face recognition algorithms, such as eigenfaces [4] or Fisherfaces [5]. The first one is local appearance modeling, 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. The other important point is data independent bases, which eliminates the need of subspace computation. In addition, DCT provides fast feature extraction, which facilitates real-time processing. Moreover, this algorithm achieved the highest correct recognition rates in the CLEAR 2007 evaluations [17], and found to perform robustly under real-world conditions [18].

The LAFR approach can be summarized as follows: A detected and registered face image is divided into blocks of 8×8 pixels size. Afterwards, on each 8×8 pixels block, the DCT is performed. The obtained DCT coefficients are ordered using zig-zag scanning. From the ordered coefficients, according to the feature selection strategy, M of them are selected and normalized resulting in an M -dimensional local feature vector. 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 the L1 norm as the distance metric [16].



Fig. 3. Sample images from the AR face database. The image on the left is a sample training image. The one in the middle is a sample image with upper face occlusion and the one on the right is a sample image with lower face occlusion.

IV. EXPERIMENTAL RESULTS

The proposed face registration algorithm is evaluated extensively under different conditions using the AR face database [13]. 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 against upper face occlusion face images with annotations “8/21: wearing sunglasses”, and for testing against lower face occlusion the ones with annotations “11/24: wearing scarf” are used, 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. In the data set, there are 110 subjects who have all these samples in both of the sessions. Sample images from the database can be seen in Fig. 3. Setups of the experiments are given in Table I.

In the experiments, face images are scaled to 64×64 pixels resolution. Ten-dimensional local feature vectors that are extracted from each 8×8 pixels block by removing the first DCT coefficient and keeping the following first ten of them are used. The selected coefficients are divided by their standard deviations and normalized to unit norm.

A. Face Recognition Without Facial Feature Localization

The experimental results of the proposed automatic face recognition system which does not need facial feature localization are presented in Table II. The achieved correct recognition rates are very low when the training images are aligned using only the manual eye center labels and the test face images are detected with a generic automatic face detector [15] and aligned using the estimated eye center positions. The results improve significantly when all the eye center combinations within the determined eye center regions are used to align the training face images. For example, the correct recognition rate increases from 32.7% to 97.3% in the *ARIsscarf* experiment and from 20% to 95.5% in the *ARIsun* experiment. As can be seen from the table, there is no significant performance difference between brute-force search and hierarchical search when the window size is $W = 11$. However, the results are lower, when the window size is $W = 9$, especially on the experiments that contain occlusion. The main reason is, due to occlusion, face detection quality is very low in these cases, which in turn causes poor eye center position estimates that cannot be covered with a smaller window size. Another interesting observation that

TABLE I
EXPERIMENTAL SETUPS

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 II

OBTAINED CORRECT RECOGNITION RATES. THE RESULTS IN THE FIRST COLUMN ARE OBTAINED USING JUST THE MANUALLY LABELED EYE CENTER POSITIONS TO ALIGN A TRAINING SAMPLE. THE ONES IN THE OTHER COLUMNS ARE ACHIEVED USING ALL THE EYE CENTER POSITION COMBINATIONS TO ALIGN THE TRAINING SAMPLE. EITHER BRUTE-FORCE OR HIERARCHICAL SEARCH IS CONDUCTED TO FIND THE BEST MATCHING ALIGNED TRAINING SAMPLE.

	Est. labels only	Brute-force search	Hierar. search $W = 11$	Hierar. search $W = 9$
AR1scarf	32.7%	97.3%	97.3%	94.6%
AR1sun	20.0%	95.5%	95.5%	90.9%
ARneutral	42.7%	100%	100%	97.3%
AR2scarf	40.0%	90.0%	89.1%	84.6%
AR2sun	13.6%	93.6%	93.6%	79.1%

can be derived from the table is that, very high correct classification rates are obtained against occlusion problem. Especially eye region occlusion is known to be one of the biggest challenges in face recognition. The obtained results imply that mainly the erroneous feature localization, thus imprecise face alignment causes the poor performance in the case of eye region occlusion. It is also intriguing to observe that the correct recognition rate obtained in the *AR2scarf* experiment is lower than the one obtained in the *AR2sun* experiments, although lower face occlusion is known to be an easier problem than the upper face occlusion. The reason can be the textured surface of the scarfs which might affect the classification decision more than the black sunglasses. It should be also considered that, as already shown, the main issue with the upper face occlusion is the misalignment and once it is handled, very high performance can be reached. Note that the achieved correct identification rates are significantly higher than the ones presented in the literature [10], [19], [20], [21], [22], [23].

B. Robust Face Recognition against Facial Feature Localization Errors

Although the proposed system is free from feature localization, it is still possible to integrate a feature detector to it. In this case, instead of using the estimated eye center positions, the output of the facial feature detector is used as the initial point of the optimization process. In the scaled faces, the distance between the eyes is 27 pixels. Conducting search within ± 5 pixels in the case of window size $W = 11$

provides insensitivity to the localization errors of up to 18% of the interocular distance, whereas conducting search within ± 4 pixels in the case of window size $W = 9$ provides insensitivity to the localization errors of up to 14% of the interocular distance.

To analyze the contribution of the proposed method to the performance of the face recognition system that uses a separate feature detection step, several experiments are conducted by using the manually labeled eye center positions of the test face image as the localization output of an automatic feature detector and adding different levels of noise to them to imitate registration errors. In the experiments, training images are registered with respect to the original, manually annotated eye center labels, whereas equally distributed random noise, ranging from 1% to 15% of the distance between the eyes, are added to the manual eye center labels of the test images. For each noise level a separate classification is done. Correct recognition rates obtained by brute-force and hierarchical search with window size $W = 11$ are plotted. In these experiments, we have also conducted hierarchical search with window size $W = 9$, which has been found to perform similarly to the hierarchical search with window size $W = 11$, except at the noise level of 15% of the distance between the eyes, where a slight decrease in performance has been observed.

In Fig. 4, the correct recognition rates obtained on the AR scarf experiments are depicted. The *AR1scarf* and the *AR2scarf* plots correspond to the results attained using just the provided eye center positions. The plots with the suffix “bf” correspond to using multiple eye center position combinations and doing brute-force search, and the ones with the suffix “hs” correspond to using multiple eye center position combinations and doing hierarchical search. As can be observed even without adding any errors and using the provided manual eye center labels (the point “0” in the x-axis), the proposed method improves the correct recognition rates significantly. On *AR1scarf*, the performance increases from 91.8% to 97.3% and on *AR2scarf*, it increases from 83.6% to 90.0%. Both brute-force and hierarchical search provided the same results at this level. As can be seen, as the error increases, the performance deteriorates, however, with the proposed method the achieved correct recognition rates stay consistent with respect to different error levels. No significant performance difference is observed between doing brute-force search and hierarchical search.

The correct recognition rates obtained on the AR sun

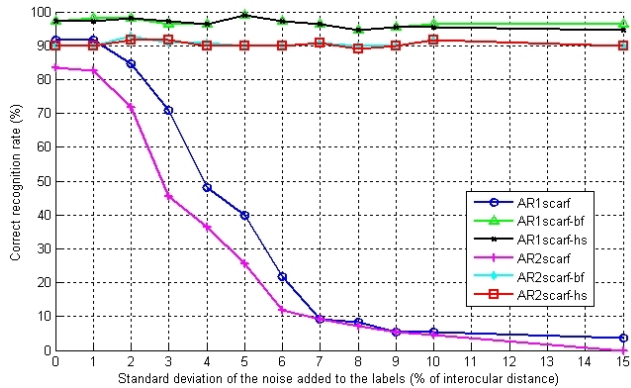


Fig. 4. Performance with respect to localization errors. The plots without suffix correspond to using only provided noisy labels, the ones with the suffix “bf” correspond to using multiple eye center position combinations and doing brute-force search, and the ones with the suffix “hs” correspond to using multiple eye center position combinations and doing hierarchical search. The proposed approach provides stable results over localization errors.

experiments are plotted in Fig. 5. The improvement in correct classification rates provided by the proposed approach is even more remarkable in these experiments. The correct recognition rate is 38.2% on the *AR1sun* and 37.3% on the *AR2sun* experiments when only the manual eye center labels are used. They become 97.3% and 95.5%, respectively, when all the eye center position combinations are utilized within the determined eye region. This outcome is not surprising, since it is not possible to precisely label the actual eye centers even manually due to occlusion caused by sunglasses which leads to misalignment. Again, correct recognition rates remain stable with respect to varying error levels. However, this time, it decreases when 15% of the interocular distance is added to the manual eye center labels as noise. As stated before, in these experiments, manual eye center labels are assumed to be precise and the errors are induced to these labels in order to imitate the localization errors. Nevertheless, these labels are not precise in the case of wearing sunglasses. Because of this reason, when the eye center positions are modified by 15% of the interocular distance, depending on how precise the manual label is, on some test images the modification could be higher than 18% of the interocular distance with respect to the actual eye center position, which is the upper limit of registration error that the proposed system can tolerate.

Again note that the achieved correct recognition rates are very high compared to the ones presented in the literature that are attained on the same database [10], [19], [20], [21], [22], [23]. Moreover, the system is also very fast. For instance, classifying a single image in the AR database takes only 10.8ms in a Pentium IV, 3GHz PC, which is significantly faster than the approach presented in [23], in which, it is stated that, 75 seconds is required per test image on a PowerMac G5.

The results of using only manual labels and the proposed alignment approach are given in Table III. As mentioned before, the correct recognition rates obtained with the proposed

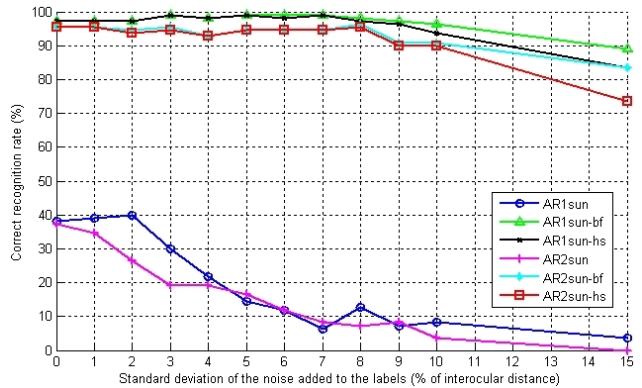


Fig. 5. Performance with respect to localization errors.

TABLE III

OBTAINED CORRECT RECOGNITION RATES. THE RESULTS IN THE FIRST COLUMN ARE OBTAINED USING JUST THE MANUALLY LABELED EYE CENTER POSITIONS TO ALIGN A TRAINING SAMPLE. THE ONES IN THE OTHER COLUMNS ARE ACHIEVED USING ALL THE EYE CENTER POSITION COMBINATIONS TO ALIGN THE TRAINING SAMPLE. EITHER BRUTE-FORCE OR HIERARCHICAL SEARCH IS CONDUCTED TO FIND THE BEST MATCHING ALIGNED TRAINING SAMPLE.

	Manual labels only	Brute-force search	Hierar. search $W = 11$	Hierar. search $W = 9$
AR1scarf	91.8%	97.3%	97.3%	98.2%
AR1sun	38.2%	97.3%	97.3%	99.1%
ARneutral	92.7%	100%	100%	99.1%
AR2scarf	83.6%	90%	90%	88.2%
AR2sun	37.3%	95.5%	95.5%	94.6%

alignment approach is superior to the ones obtained using the manual labels. No significant performance difference is observed between doing hierarchical search with $W = 11$ and $W = 9$. It is the first time in the literature that higher performance is achieved by using automatic alignment compared to using manual alignment performed by utilizing manually labeled facial feature points.

It is intriguing to investigate whether the stable performance with respect to misalignment and significant performance increase in the case of occlusion are solely provided by the proposed face registration approach or its integration to the local appearance-based face recognition using DCT [16]. Therefore, we integrated the proposed registration algorithm to eigenfaces approach [4] and conducted the same experiments. Mahalanobis cosine (MAHCOS) distance metric is used in nearest neighbor classification. The obtained results can be seen from Table IV. It is clear that eigenfaces algorithm cannot benefit from the proposed face alignment technique, mainly, due to its data-dependent bases.

V. CONCLUSION

In this paper, we addressed the misalignment problem in face recognition systems. Different from traditional face recognition systems, the proposed algorithm does not need an additional facial feature localization step for face registration.

TABLE IV

PERFORMANCE OF EIGENFACES ALGORITHM. THE RESULTS IN THE FIRST COLUMN ARE OBTAINED USING JUST THE MANUALLY LABELED EYE CENTER POSITIONS TO ALIGN A TRAINING SAMPLE. THE ONES IN THE SECOND COLUMN ARE ACHIEVED USING ALL THE EYE CENTER POSITION COMBINATIONS TO ALIGN THE TRAINING SAMPLE. BRUTE-FORCE SEARCH IS CONDUCTED TO FIND THE BEST MATCHING ALIGNED TRAINING SAMPLE.

	Eigenfaces manual—	Eigenfaces auto—
AR1scarf	28.2%	23.6%
AR1sun	17.3%	26.4%
AR2scarf	17.3%	15.5%
AR2sun	17.3%	23.6%

It implicitly performs feature localization at the classification step. There are four main findings from this study.

- It has been shown that, facial feature localization step can be eliminated from the face recognition systems and the alignment can be performed by directly aiming at minimizing the closest classification distance. Very high correct recognition rates have been achieved on the AR face database [13].
- It has been shown that the proposed registration approach performs even better than doing registration with manual labels. For instance, on the AR face database [13], against lower facial occlusion, the obtained result when the test images are aligned using the manual labels was 91.8%, while with the proposed registration approach it has become 97.3%.
- It has been found that the main problem with the upper face occlusion is due to registration errors, and not the occlusion itself. Due to the sunglasses, the eye center points that are widely used for face alignment can not be reliably labeled even manually. When only the manual labels are used to align the test images, the achieved correct recognition rate against upper facial occlusion with sunglasses is 38.2% on the AR face database [13]. The performance jumps to 97.3% with the proposed registration approach.
- The optimization procedure integrated to the classification step makes the face recognition system insensitive to the facial feature localization errors. The algorithm can tolerate up to 18% of the interocular distance as localization error, and up to this point it provides stable performance.

It is important to note that although previous studies [10], [11] also tried to utilize additional samples, they were not able to provide stable performance with respect to misalignment. The stable performance and significant performance increase in the case of occlusion is only possible with the integration of the proposed face alignment approach to the local appearance-based face recognition using DCT [16].

REFERENCES

- [1] S. Arca, P. Campadelli, and R. Lanzarotti, "A face recognition system based on automatically determined facial fiducial points," *Pattern Recognition*, vol. 39, no. 3, pp. 432–443, 2006.
- [2] A. Lemieux and M. Parizeau, "Experiments on eigenfaces robustness," in *Proc. Int'l. Conf. on Pattern Recognition*, vol. 1, 2002, pp. 421–424.
- [3] E. Rentzeperis, A. Stergiou, A. Pneumatikakis, and L. Polymenakos, "Impact of face registration errors on recognition," in *Artificial Intelligence Applications and Innovations*. Springer, 2006, pp. 187–194.
- [4] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class-specific linear projection," *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.
- [6] L. Wiskott, J. M. Fellous, N. Kruger, and C. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 19, no. 7, pp. 775–779, 1997.
- [7] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 27, no. 3, pp. 328–340, 2005.
- [8] C. Xie, B. V. K. V. Kumar, S. Palanivel, and B. Yegnanarayana, "A still-to-video face verification system using advanced correlation filters," in *Proceedings of Intl. Conference on Biometric Authentication*, 2004, pp. 102–108.
- [9] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, 1994, pp. 138–142.
- [10] A. M. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class," *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 24, no. 6, pp. 748–763, 2002.
- [11] S. Shan, Y. Chang, W. Gao, B. Cao, and P. Yang, "Curse of misalignment in face recognition: Problem and a novel mis-alignment learning solution," in *Proc. of the 6th IEEE Intl. Conf. on Automatic Face and Gesture Recognition (FGR'04)*, 2004, pp. 314–320.
- [12] H. Wang, S. Yan, T. Huang, J. Liu, and X. Tang, "Misalignment-robust face recognition," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2008)*, 2008, pp. 1–6.
- [13] A. M. Martinez and R. Benavente, "The AR face database," CVC, Tech. Rep. 24, 1998.
- [14] P. J. Phillips and et al., "Overview of the Face Recognition Grand Challenge," in *Proc. 2005 IEEE Conf. on Computer Vision and Pattern Recognition*.
- [15] Intel Corporation, "Open source computer vision library (OpenCV)," 2008, last visit: Nov. 2008. [Online]. Available: <http://www.intel.com/technology/computing/opencv/index.htm>
- [16] H. K. Ekenel and R. Stiefelwagen, "Analysis of local appearance-based face recognition: Effects of feature selection and feature normalization," in *Conf. on Computer Vision and Pattern Recognition Workshop*, 2006, p. 34.
- [17] R. Stiefelwagen, K. Bernardin, R. T. R. Bowers, M. Michel, and J. Garofolo, "The CLEAR 2007 evaluation," in *Proc. of the Int'l. Evaluation Workshops CLEAR 2007 and RT 2007*, 2007, pp. 3–34.
- [18] J. Stallkamp, H. K. Ekenel, and R. Stiefelwagen, "Video-based face recognition on real-world data," in *Proc. of IEEE Intl. Conf. on Computer Vision, ICCV*, 2007, pp. 1–8.
- [19] X. Tan, S. Chen, Z. H. Zhou, and F. Zhang, "Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k -NN ensemble," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 875–886, 2005.
- [20] B. G. Park, K. M. Lee, and S. U. Lee, "Face recognition using face-arg matching," *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 27, no. 12, pp. 1982–1988, 2005.
- [21] S. Fidler, D. Skočaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling," *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 28, no. 3, pp. 337–350, 2006.
- [22] H. Jia and A. M. Martinez, "Face recognition with occlusions in the training and testing sets," in *Proc. of IEEE Int'l. Conf. on Automatic Face and Gesture Recognition*, 2008.
- [23] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 31, no. 210–227, pp. 775–779, 2009.