Eye Perturbation Approach for Robust Recognition of Inaccurately Aligned Faces

Jaesik Min, Kevin W. Bowyer, and Patrick J. Flynn

Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

Abstract. Extraction of normalized face from input images is an essential preprocessing step of many face recognition algorithms. Typical face extraction algorithms make use of the locations of facial features, such as the center of eyes, that are marked either manually or automatically. It is not guaranteed, however, that we always obtain the exact or optimal locations of the eye centers, and using inaccurate landmark locations, and consequently poorly registered faces, is one of the main causes of performance degradation in appearance-based face recognition. Moreover, in some applications, it is hard to verify the correctness of the face extraction for every query image. For improved performance and robustness to the eye location variation, we propose an eye perturbation approach that generates multiple face extractions from a query image by using the perturbed eye locations centered at the initial eye locations. The extracted faces are then matched against the enrolled gallery set to produce individual similarity scores. Final decisions can be made by using various committee methods – nearest neighbor, maximum vote, *etc.*– of combining the results of individual classifiers. We conclude that the proposed eye perturbation approach with nearest neighbor classification improves recognition performance and makes existing face recognition algorithms robust to eye localization errors.

1 Introduction

Many face recognition methodologies require, as an essential preprocessing step, the extraction of a normalized face region from the input image. In many appearance-based face recognition approaches, the face extraction is performed based on the locations of facial landmarks, such as eyes, nose, or mouth [1]. Once the coordinates of these landmarks are given, extraction of the face can be done through the processes of image scaling, rotation, intensity normalization, and aligning to a predetermined template, *etc.* that minimizes the variations unrelated to the identity.

The most prominent facial landmarks in 2D face images are the eyes [2], whereas it is the nose in 3D (depth) face images [3]. The locations of eye centers can be obtained either manually or automatically by using eye detection algorithms [4]. Often, however, the detected eye locations are unreliable; they are inaccurate and inconsistent across eye detectors. This causes sub-optimal face extraction, and consequently degrades recognition performance even with a good algorithm and images of well-posed faces [5]. In this paper we first investigates the effect of the accuracy of eye locations.

Minimizing the errors at the stage of localization is desired for this problem, but has a limit. An alternative solution is to take the existence of localization errors for granted,

T. Kanade, A. Jain, and N.K. Ratha (Eds.): AVBPA 2005, LNCS 3546, pp. 41–50, 2005.

and to design a recognition algorithm that is robust to the localization variation. In this paper we propose to produce multiple eye locations perturbed from the initial locations of both eyes and then use the extracted faces from these eye locations. We tested two representative face recognition algorithms, PCA and FaceIt, on a large number of face extractions that are generated from various sampling of eye locations. Then we compared the results of eye perturbation to the baseline.

The remaining sections of this paper are organized as follows. In Section 2, a number of related works are investigated. Sections 3 to 5 describe how we designed the experiments on eye perturbation and committee and discusses the effect of these factors on the performance. Section 6 shows compared results of the experiments. Section 7 summarizes our work and introduces topics to be addressed in future work.

2 Previous Works

The importance of eye localization as a preprocessing module in a face recognition system has been addressed by many researchers. Marques *et al.* [6] investigated the effect of eye position on a PCA-based face recognition algorithm. They used a total of 8 images and showed the sensitivity of the algorithm to the eye location deviations along various directions. As mentioned in their work, even the eye positions that are manually selected – or at least inspected – by human operators are unreliable and tend to deviate from a definition of the geometric eye center.

The role of eye locations in achieving high performance in face recognition systems received special focus in the paper by Riopka *et al.* [2]. They evaluated the effect of eye location accuracy through experiments of 3 different face recognition algorithms, that is, Principal Component Analysis (PCA), Elastic Bunch Graph Matching (EBGM), and FaceIt, on 1024 images from FERET database [7] by generating $17\times17=289$ perturbations of eye locations from the original locations and compared the recognition results. They first used ideal image data – that is, used the same image set for both gallery and probe sets – to measure the pure effect of eye perturbation. Then they applied the same perturbation to more realistic images. They report that using real image data did not degrade the performance drastically when the same eye perturbation is applied.

Some researchers have proposed solutions to the inaccurate localization problem. In the paper by Martinez [8], the gallery is augmented by perturbation and modeled by Gaussian Mixture Models (GMM). Shan *et al.* [9] defined robustness to misalignment in their paper and observed the effects of misalignment. They also proposed an enhanced Linear Discriminant Analysis (LDA) algorithm for face recognition that generated multiple ($9\times9=81$) virtual samples from each original training image by perturbation.

3 Experimental Design

A total of 600 subjects were selected partly from the FERET database [7] and partly from the University of Notre Dame (ND) database [10] so that their neutral expression face images are used in creating a training image set. Another 393 subjects from the ND database who participated between years of 2002 and 2003 were selected to create a test image set where each subject's earliest image is used as the gallery image and the

Fig. 1. Gallery and probe image samples of the University of Notre Dame database. Each image is either 1200×1600 or 1704×2272-pixel color image

latest image is put into the probe set (Figure 1). The elapsed time between the gallery and probe images ranges from 1 to 94 weeks, with 35 weeks on average. Both gallery and probe images are acquired under the same controlled environment, that is, lighting condition, background, and facial expression. There may exist slight and unintended pose variation and other variations over time.

As the recognition algorithms, we used two representative face recognition algorithms: PCA and FaceIt. For the PCA algorithm, we used the Version 5.0 code implemented at Colorado State University (CSU) [11]. The Mahalanobis Angle was selected as the distance metric, and no dimension reduction of the eigenspace was performed. For the FaceIt algorithm, we used the version G5, which was developed and distributed by Identix Incorporated.

The recognition performance is represented by a cumulative match characteristic (CMC) score, where CMC score at rank r is defined as the ratio of people whose correct match exists within r best matches. For example, a score of 85% at rank 1 means that 85% of people were correctly matched at the first choice. Similarly, a score of 90% at rank 3 means that 90% of people have their correct matches in the first three best matches. Therefore, a single recognition result gives different scores at different ranks, and the score at rank s is higher than or equal to the score at rank r, where $r < s$.

4 Effect of Inaccurate Localization

Previous studies [2, 9] show that PCA and LDA algorithms are sensitive to eye localization errors. Figure 2 shows the real examples of face extraction when eyes are localized by the eye locater module of the FaceIt software. Inaccurate localization yields undesirable, *e.g.*, scaled, rotated, or translated face templates. In this section, we investigate how the inaccurate localization affects the performance of algorithms. For this we set the manually marked eye locations as the ground truth and the automatically selected eye locations as the set of real-life samples, because it sounds more practical to get samples from a real eye locater rather than to add artificial random noises to the ground truth locations.

Originally, all of the 393 gallery images and 393 probe images are provided with ground truth eye locations. The probe images were also fed into the eye locater module of the FaceIt software to get the eyes localized automatically. These locations are compared to the ground truth eye locations of the same images. The difference between the

scaled scaled rotated translated

Fig. 2. Examples of poorly extracted faces due to the eye position deviation. Faces at the top row are from gallery images with ground truth eye positions and faces at the bottom are from corresponding probe images with automatically marked eye positions

manual and the automated eye locations is 10.7 pixels on average, with standard deviation of 5.7 pixels, while the average distance between two eye centers in the ground truth is 268.8 pixels.

We ran three face recognition algorithms at hand on this gallery set (with ground truth) and probe set (automatic markings). Figure 3 shows how the performance of each algorithm degrades with the inaccurate eye locations. As expected, the PCA algorithm degrades abruptly, confirming that it is highly dependent on the localization accuracy. The FaceIt and EBGM algorithms turned out to be relatively tolerant to the inaccurate localizations; the performance also degrades, but the amount of degradation is negligible. We do not know what FaceIt does to handle this problem, and EBGM adjusts itself to some degree. Similar experiments with FaceIt performed in [2] showed large degradation with the "weathered" image set. In the next session we propose a method of augmenting the probe data to solve the problem caused by misalignment.

5 Eye Perturbations

Using large and representative samples per class is the best way to assure better classification, but it is not always feasible [8]. Generating multiple versions of face templates from a limited number of originals, thus augmenting the dataset, is one promising solution, as in [8, 9]. To solve the problem of inaccurate eye localization as discussed in Section 4, we propose to augment the probe images by perturbing the initial eye locations.

In real-life applications, the gallery set resides in the database, thus its quality and metadata are under strict control. In contrast, the probe images usually are transient and its quality (along with that of metadata) is less controlled. Thus, it is more likely the probe images have bad eye localization than the gallery images do. Therefore, instead of augmenting the gallery set as in [8, 9], we propose to augment the probe images. By keeping the gallery set and augmenting the probe images, the face recognition system becomes more flexible in that the degree of dataset augmentation is easily adjustable accordingly; there is no need of rebuilding or remodeling of the system.

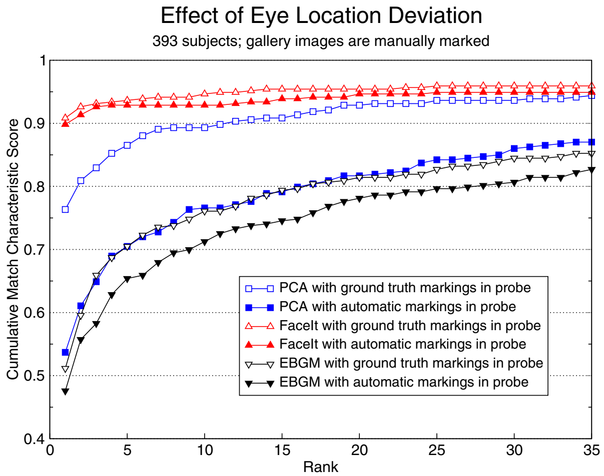


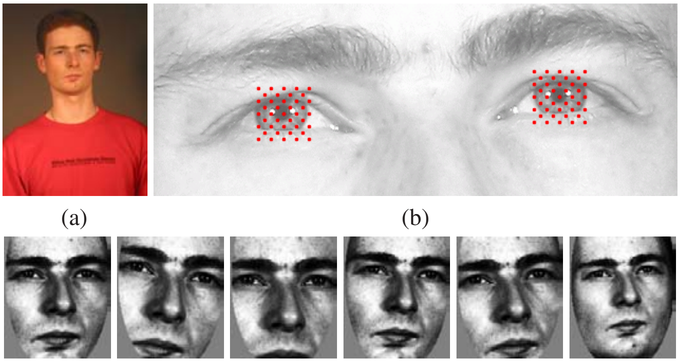
Fig. 3. The performance degradation due to inaccurate eye localization. The CMC at rank 1 of PCA algorithm dropped 76.3 % to 53.5 %. The EBGM (51.1 % to 47.6 %) and FaceIt (90.8 % to 89.8 %) were less affected by the deviation

All of 786 images used in our experiments are accompanied by the ground truth facial landmarks, which are marked by a number of different human operators and are highly reliable in spite of the existence of slight variations across operators and over time. The images also are provided with the machine-selected eye locations. For each original query image, we generate multiple normalized faces by perturbing the initial – either ground truth or machine-selected – eye locations (Figure 4). The sampling window size is set to 49×49 pixels so that it covers an area slightly wider than the iris. We sample 41 uniformly distributed locations for each eye, a total of 41×41=1681 pairs of eye locations, and thus generate the same number of normalized faces for each query image.

Each normalized face probe matches against the gallery set and produces distance measures to each of the 393 gallery images. So for each query image we will have 1681 individual classification results. A number of committee schemes to combine these results are available, such as nearest neighbor, k-nearest neighbors, weighted sum, maximum vote, *etc.* So far we tested the nearest neighbor and maximum votes. In the nearest neighbor (NN) scheme, we simply select the pair of probe and gallery with the minimum distance – or the highest similarity score in FaceIt terminology. In the maximum vote scheme, the gallery image that gets the maximum number of NN selections from 1681 individual normalized face probes is finally selected.

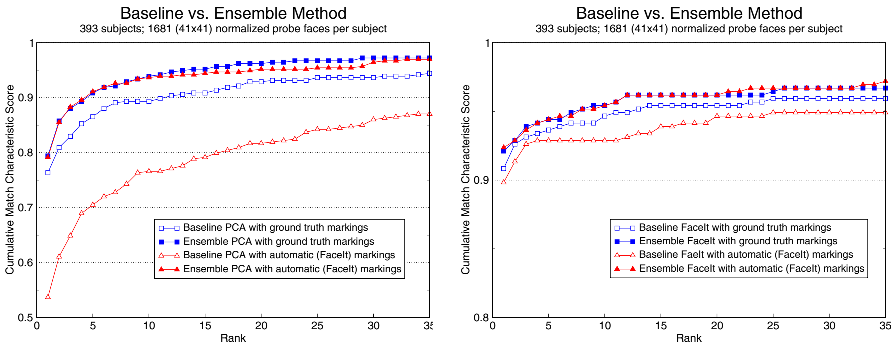
6 Results and Discussion

We compared the NN ensemble method to the baseline on the PCA and the FaceIt algorithms, where both ground truth and machine-selected eye markings are provided (Fig. 5 (a)). The NN ensemble PCA algorithm scored 79.4 % rank-1 CMC, marginally improved from that of the baseline PCA, 76.3 %. The improvement achieved by the



(a) (b) (c)

Fig. 4. An example of multiple generation of normalized faces from a probe image. Given an original image (a), possibly with inaccurate eye locations, 41 sampling locations centered at the initial eye locations are selected for each eye as illustrated in (b). Six out of 1681 normalized faces are shown in (c)

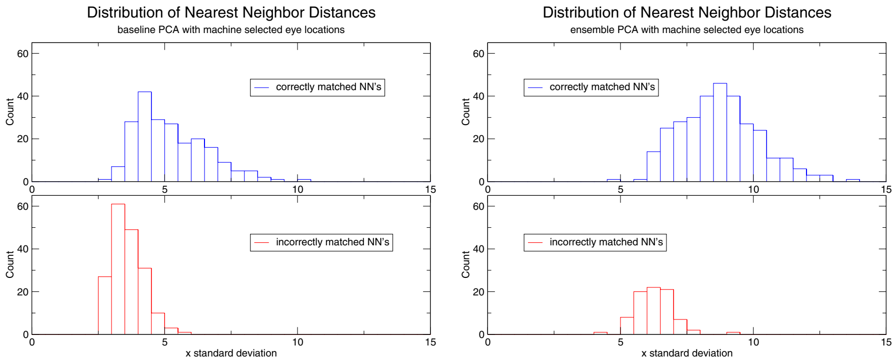


(a) PCA (b) FaceIt

Fig. 5. Comparison between the baseline and ensemble methods. The ensemble methods achieved both improvement (significant or marginal) and stability in performance

NN ensemble is just over 3 %, but considering that the baseline performance was with the ground truth, it promises greater improvement with machine-selected eye locations. The experiment of the ensemble PCA algorithm with machine-selected eye locations reached 79.1 % rank-1 CMC, which is a huge improvement from the baseline performance of 53.7 %. The comparison of baseline and ensemble FaceIt is shown in Fig. 5 (b). The baseline performance of FaceIt is already high enough, but we still observe marginal improvements, and the amount of improvement is a little higher in case of machine-selected eye locations, which was expected. The overall CMC curves shown in Figure 5 indicate that the ensemble method also achieved stability in performance as well as improvement, that is, we observe only negligible difference in performance between the ground truth and automatic markings.

At this point we need to analyze the mechanism of the maximum vote scheme, which yields low performance. The maximum-vote ensemble method was also applied



(a) baseline PCA (b) ensemble PCA

Fig. 6. The distribution of nearest neighbor distances in the (a) baseline and (b) ensemble PCA with the automatically marked eye locations. Each of the 393 probes have 393 distance measures to the gallery images, and this plot shows how the 393 NN distance outlies in the overall distribution of probe-to-gallery distances

with the PCA algorithm, but it achieved lower performance (63.6 % rank-1 CMC) than that of the baseline. In general, the NN pair of a probe and a gallery image is the extreme outlier in the distribution of distances between probe and gallery images. We investigated how far the NN distance lies in the distance distribution. In the baseline experiment the correctly matched NN distances lie at, on average, 5.2 standard deviation of the distance distribution, and the incorrectly matched NN distances lie at 3.6 standard deviation (Figure 6 (a)). In both cases, the NN distances are the extreme outliers in the distribution whose p-values are less than 0.001. This extremity of the NN distance gets further (Figure 6 (b)) in the ensemble scheme because it produces a better (or equal at least) NN distance and adds a huge amount of mediocre distances. This explains the poor performance achieved by the maximum vote committee method, where the newly produced NN distance just casts one vote equally as the other 1680 distances do. Therefore, hereinafter we discard the maximum vote committee scheme and focus on the NN scheme only.

Figure 7 shows two examples of successful NN match after the eye perturbation. At the top row, the original probe in the baseline was matched to a wrong gallery image, and the correct gallery image scored rank of 150; after the eye perturbation, one of the perturbed probes was matched to the correct gallery image. At the bottom row of the figure, which is the case where the machine-selected eye locations were provided, the rank score has jumped from 131 to 1.

However, augmenting the dataset not always improve the performance. It is possible that some of the enlarged data may match to wrong gallery images with smaller distances than that of correct match. In our experiments it actually happened (Figure 8), but the rank change is relatively small. The count and amount of performance improvement and degradation are summarized in Table 1. In the PCA case with ground truth, the number of instances of improvement is not much more than that of degradation, but the average amount of rank change is larger, which gives overall improvement. In the PCA case with machine-selected eye locations, both the number and the amount of rank

(a) (b) (c) (d)

Fig. 7. Successful cases of eye perturbation. The probe with initial eye localization (a) is matched to a wrong gallery image (b). After perturbation, a new probe image (c) is successfully matched to the correct gallery image (d). It is shown that the probe image (a) and the gallery image (d) are not well aligned

(a) (b) (c) (d)

Fig. 8. Examples of degradation after eye perturbation. The probe with initial eye localization (a) is matched to a correct gallery image (b). After perturbation, a new probe (c) picked up a wrong match (d) that has smaller distance

change is large, which explains the big jump in the CMC curve in Figure 5. The FaceIt rank results have similar patterns, although less obvious.

There also exist cases where the proposed method cannot be the solution. The subject in Figure 9, for example, has significant pose change between the gallery and probe images. Neither PCA nor FaceIt succeeded in matching this subject correctly both in the baseline and in the ensemble method because the problem here comes from the pose angle rather than from the localization accuracy.

7 Conclusions and Future Works

In this paper we showed the effect of inaccurate eye localization on the performance of face recognition and proposed a method that is robust to the effect. We first investigated the impact of eye localization accuracy through experiments with two sets of realistic localization data; a set of manual markings, which is used as the ground truth, and another set automatically marked by a commercial software, which served as the devi-

Table 1. The rank change between the baseline and the NN ensemble methods

	PCA				FaceIt			
	GT		Auto		GT		Auto	
	Count	Amount	Count	Amount	Count	Amount	Count	Amount
Improved	**69**	**36.4**	**160**	**44.2**	**22**	**16.7**	**31**	**53.6**
Degraded	**51**	**7.9**	**28**	**16.3**	**13**	**19.2**	**10**	**13.3**
Unchanged	273		205		358		352	

(a) (b) (c) (d) (e)

Fig. 9. A failed match after eye sampling. The various extractions of the probe images, (a) and (c), could not be matched to the correct gallery image, (e), because the pose difference between the original gallery and probe images is significant. Image (c) and (d) are the nearest neighbor pair after the eye perturbation

ation from the ground truth. By using large sets of images with substantial time lapse between the gallery and probe images, and by using real-life outputs of eye localization, we showed that, for some face recognition algorithms, the accuracy of eye localization is critical to the recognition performance.

Based on the baseline experimental results, we proposed an eye perturbation approach to make existing face recognition algorithms robust to the eye localization variation. A number of experiments with ground truth and machine-selected eye locations showed that achieving both improvement and robustness was successful.

It will be worth investigation to extend this experiments with image sets of larger variety. As mentioned in [2], the inaccurate eye localization may have the greatest impact on controlled pairs of gallery and probe images; using pairs of different conditions in the probe images – *e.g.*, uncontrolled probe images against controlled gallery – might attenuate the effect of inaccurate localization.

Currently the increased computational cost is the main problem of the proposed approach. We used a full-scale eye perturbation for a thorough investigation, but a smaller and sparser sampling may be enough for the intended purpose. Alternatively, the degree of perturbation may be parameterized so that the degree can be adjustable. We also plan to design an intelligent decision algorithm by modeling the distribution of NN distances as shown in Figure 6, so that it can decide the necessity and degree of the eye perturbation, methods of combining individual classifications, *etc.*

Acknowledgments

This work is supported by National Science Foundation grant EIA 01-20839 and Department of Justice grant 2004-DD-BX-1224.

References

1. Chellappa, R., Wilson, C., Sirohey, S.: Human and machine recognition of faces: A survey. Proceedings of IEEE **83** (1995) 705–740
2. Riopka, T., Boult, T.: The eyes have it. In: Proc. of the ACM SIGMM workshop on biometric methods and applications. (2003) 9–16
3. Chang, K., Bowyer, K., Flynn, J.: Face recognition using 2D and 3D facial data. In: ACM Workshop on Multimodal User Authentication. (2003) 25–32
4. Hsu, R., Abdel-Mottaleb, M., Jain, A.: Face detection in color images. IEEE Trans. Pattern Anal. and Mach. Intel. **24** (2002) 696–706
5. Zhao, W.: Improving the robustness of face recognition. In: Audio- and Video-Based Biometric Person Authentication. (1999) 78–83
6. Marques, J., Orlans, N., Piszcz, A.: Effects of eye position on eigenface-based face recognition scoring. Technical Report, MITRE Corporation (2000)
7. Phillips, P., Moon, H., Rizvi, S., Rauss, P.: The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. and Mach. Intel. **22** (2000) 1090–1104
8. Martinez, A.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Trans. Pattern Anal. and Mach. Intel. **24** (2002) 748–763
9. Shan, S., Chang, Y., Gao, W., Cao, B., Yang, P.: Curse of mis-alignment in face recognition: Problem and a novel mis-alignment learning solution. In: International Conference on Automatic Face and Gesture Recognition. (2004) 314–320
10. Min, J., Flynn, P., Bowyer, K.: Assessment of time dependency in face recognition. TR-04-12, University of Notre Dame (2004)
11. Beveridge, R., Draper, B.: Evaluation of face recognition algorithms (release version 5.0) (URL: http://www.cs.colostate.edu/evalfacerec/index.html)