Classification Enhancement via Biometric Pattern Perturbation*

Terry Riopka[1] and Terrance Boult[2]

[1] Lehigh University, Dept. of Computer Science and Engineering
Bethlehem, PA 18015 U.S.A.
riopka@kantbelievemyeyes.com

[2] University of Colorado at Colorado Springs, Computer Science Dept.
Colorado Springs, CO 80933 U.S.A.
tboult@cs.uccs.edu

Abstract. This paper presents a novel technique for improving face recognition performance by predicting system failure, and, if necessary, perturbing eye coordinate inputs and repredicting failure as a means of selecting the optimal perturbation for correct classification. This relies on a method that can accurately identify patterns that can lead to more accurate classification, without modifying the classification algorithm itself. To this end, a neural network is used to learn 'good' and 'bad' wavelet transforms of similarity score distributions from an analysis of the gallery. In production, face images with a high likelihood of having been incorrectly matched are reprocessed using perturbed eye coordinate inputs, and the best results used to "correct" the initial results. The overall approach suggest a more general approach involving the use of input perturbations for increasing classifier performance in general. Results for both commercial and research face-based biometrics are presented using both simulated and real data. The statistically significant results show the strong potential for this to improve system performance, especially with uncooperative subjects.

1 Introduction

Face detection is a critical preprocessing step for all face recognition systems. Its ultimate purpose is to localize and extract the face region of an image (which may or may not contain one or more faces) and to prepare it for the recognition stage of a face processing engine. In general, as a face preprocessor, it must achieve this task regardless of illumination, orientation or size of the input face image. As daunting as this task is for computers, it is a task that humans appear to do rather effortlessly.

Face detection approaches can be broadly organized into two categories: feature-based approaches [1], and image-based approaches [2]. The former relies primarily on he extraction of low level features incorporating face knowledge explicitly, while the latter treats the face as a pattern that can be learned from the two-dimensional

*T. Kanade, A. Jain, and N.K. Ratha (eds.): AVBPA 2005, LNCS 3546, pp. 850-859, 2005. ©Sprinter-Verlag Berlin Heiderlberg 2005.

image array, incorporating face knowledge implicitly. However, regardless of the approach, the result of face detection must enable some method for face registration, in order to maximize the effectiveness of the recognition stage of the face processor. In all cases, this relies on the accurate determination of fiducial marks on the face, ultimately needed for scaling and normalization.

Symmetry of the eyes and their consistent relationship with respect to other fiducial marks on faces make them extremely useful for parameterizing and normalizing geometric features of the face. Because eye separation does not change significantly with facial expression, nor with up and down movements of the face, eye separation distance is often used for face normalization. Nose distance, another feature often extracted, is relatively constant with respect to side to side movements of the face and also depends on accurate eye localization. In addition, orientation of the line between the eyes is often used to correct for pose variations. Lastly, eyes are essentially unaffected by other facial features like beards and mustaches, making them invaluable features to most face recognition systems. As a result, eye localization is often the critical thread connecting face detection and face recognition algorithms, regardless of the underlying details of either algorithm.

Previous studies have emphasized the critical importance of eye localization and have demonstrated the dramatic effect poor eye localization can have on face recognition [3][4]. Given that the accuracy of eye localization has an effect on face recognition performance, this paper seeks to address the following research question: **can we observe the effect that input eye perturbations have on an arbitrary recognition algorithm for a given face gallery, and use that information to improve classification performance?** The goal of this paper is to predict classification failure and, in instances in which it is expected to occur, use a failure prediction module to select an alternative eye location (perturbation) that has the greatest chance of yielding a correct classification, thus improving overall system performance.

The paper is organized as follows. A description of the method used to identify candidate face images for eye input perturbation is presented. Next, statistical results of simulated experiments explore the costs/benefits of our technique. The technique is also applied to a set of "real-world" face images to show the utility of the approach. Finally, we conclude with a discussion of the results and comment on the viability of a general approach to improving pattern classification using perturbations of critical input data.

2 Failure in the Context of Face Recognition

All face classifiers ultimately yield some sort of similarity score for an input image against all images in the face gallery. Typically, the scores are ranked to determine the most likely set of matching face images. The definition of "failure" in the context of face recognition typically depends on the application. For example, in identity verification, a serious failure occurs whenever a face not in the database is matched by the system*, i.e.* there is a false positive. In this case, the input face image matches an image in the database with a similarity score that is above a certain threshold. The decision of the system is based entirely on a comparison between two images, to determine whether the person is who the person claims to be.

In identification, the application of interest in this paper, a known or unknown individual is matched against all of the face images in the database, and a set of ranked potential matches is returned. In this case, the definition of failure is more complex. If the person is in the database, failure occurs if too many face images different from that person are ranked higher than the face image of that person in the database. Here, "too many" depends on the criteria of the system and how the results are interpreted. If the person is not in the database, it becomes problematical to determine whether or not the face is in the database based on ranking alone.

We postulate that the relationship between the similarity scores of the matched images (more specifically, the shape of their distribution) contains valuable information that can yield insight into the likelihood that a given match will lead to a correct classification. For example, intuitively, if all top ranked images have very close similarity scores, we might tend to believe there is a low probability that the top ranked image is the correct match. On the other hand, if the top ranked image has a similarity score that is significantly higher than all of the rest, we might tend to believe there is a high probability that the top ranked image is the correct match. In the former case, the distribution of sorted scores may be broad and flat, while in the latter case, narrow and peaked. Note that the criteria for "closeness" of similarity scores also depends on the characteristics of the particular recognition algorithm, since (usually) similarity score is not a metric.

In this paper, we use a machine learning approach to learn the characteristics of "good"' and "bad" similarity score distributions, given a particular recognition algorithm, a specific gallery of images, and various degrees of eye location error. "Good" similarity score distributions are those that result in a correct ID match (rank 1), where each individual (regardless of the number of images in the gallery) has a unique ID. "Bad" distributions are all others.

We make the general assumption that input eye locations are primarily responsible for classification failure as supported by [3]. Using our failure prediction model, we identify images that are likely to be classified incorrectly and then re-process those images using a limited set of perturbed input eye coordinates to yield new similarity score distributions. For each such image, the distribution most likely to yield a correct classification is identified and used to obtain a modified classification.

3 Face Algorithms

Two different face recognition algorithms were used in all of the following experiments: Elastic Bunch Graph Matching (EBGM)[5] and FaceIt, a commercial application based on an LFA algorithm [6]. The EBGM algorithm was provided by the Colorado State University (CSU) Face Identification Evaluation System (Version 5.0) [7]. FaceIt was implemented using programs built from a software development kit licensed from Identix Inc. The reader is referred to the relevant publications for details.

4 Learning Similarity Score Distributions

In order to learn similarity score distributions, a sample of “good” and “bad” similarity score distributions was required. If the intent were to learn "good" and "bad" similarity score distributions for face images *in general*, one might be inclined to train on similarity score distributions from a large set of "real" images of individuals in a given gallery. From an operational perspective and excluding synthetically altered gallery images, this would require considerable data collection and ground truth. However, the very specific intent here is to predict the behavior of a given algorithm on a given gallery with respect to input eye perturbations and to enable the recognition of potential instances where incorrect eye localization can result in misclassification. Generating the perturbation data is quite straightforward. Given some basic training/testing sets, one simply forces the eye locations to different positions and reprocesses the images.

As was shown in previous research, the behavior with respect to input eye perturbations of a number of face recognition algorithms on degraded images, seems to be quite similar to their behavior on clean, gallery images [3], only slightly smoother. Consequently, the training set in this instance involved only the similarity score distributions obtained by perturbing input eye coordinates of gallery images. The prediction module therefore learns the sensitivity of the algorithm to eye localization error in the context of the gallery for which classification improvement is desired, which we later apply, with good success, to images in the field.

4.1 Preprocessing

The images used to obtain training data consisted of a gallery of 256 individuals, each with four different frontal view poses (for a total of 1024 images) and obtained from the FERET database. The exact set of images can be obtained from the authors.

It was hypothesized that the number of poses of a given individual would affect the relevant characteristics of similarity score distributions. For example, if an individual had ten different poses in a given database, it is conceivable that all ten poses might cluster very closely in the top ranks of the similarity score distribution. On the other hand, with only one pose in the database, an individual's score might be distinctly different from all others, resulting in a similarity score distribution that is much more peaked. This suggested that a multi-resolution approach might be beneficial to extract relevant detail, which might depend on the number of poses each individual has in the database.

Recall that a wavelet basis is described by two functions (the scaling and the mother wavelet function), and a basis is obtained by translating and resizing these functions. Any signal can be represented uniquely and exactly by a linear combination of these basis functions, provided the basis functions are orthornormal. Wavelet basis functions also have a characteristic called compact support, meaning the basis functions are non-zero only on a finite interval. In contrast, the sinusoidal basis functions of the Fourier transform are infinite in extent. The compact support of the wavelet basis functions allows the wavelet transformation to efficiently represent functions or signals which have localized features.

In this application, a 4 point discrete Daubechies wavelet transform [8] was used to process the top 2k sorted similarity scores, where k is the number of poses for each individual. In this case, k=4, resulting in a total of 8 wavelet coefficents,. Reflection generated the necessary points for the function boundary. A Daubechies wavelet transform was used due to its (coarse) similarity to the distributions as well as its overlapping iterations, enabling it to pick up detail that might be missed by, say, the Haar transform.

Two additional features were also computed. The first was the next highest rank of the same ID as the top ranked image. Since only the top 2k similarity scores were observed, this number was clamped at a rank of 2k+1. Very high numbers for ranks are known, from previous experience, to be relatively unstable as predictive features. The intuition here is that the likelihood of the winner being correct is higher if the image of one of its other poses is also highly ranked.

The second feature was the number of pairs of identical Ids in the top 2k similarity scores that have a different ID from the winner. In this case, it was hypothesized that the presence of two (or more) same-ID highly ranked images in the top ranks might also have some bearing on the possibility of classification failure.

4.2 Training

Gallery images were run through each algorithm using all combinations of input eye offsets shown in figure 1, resulting in 9x9=81 runs per algorithm. Note, the same pair of eye offsets was applied to *all* of the gallery images for any given run. Random eye offsets for each individual image were not trained on, since any feasible method used in production would have to apply the same pair of offsets to the entire probe set (see section 4.3).

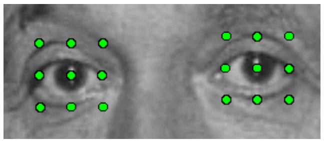

Figure 1. Eye offsets used for training.

The distance between points in the images tested was six pixels. In general, this perturbation depends on the scale of the imaged face, with the goal to select points to span the extent of the eyelids and the whites of the eyes. Similarity scores of the 8 top ranked images were stored along with the other two features discussed previously for all images. Feature vectors were generated and organized into two datasets, one for images whose rank was one (correct matches) and all others (incorrect matches).

A random sampling of 5000 out of 1024x81=82944 feature vectors was used to train a backpropogation neural network [11]. All other 77944 feature vectors were used for testing. This was done for both the FACEIT algorithm and the EBGM algorithm. Thresholds that maximized performance on the test set were fixed for all subsequent experiments and are shown in table 1, along with network architectures and performance. The neural net trained in approximately one day on a G4 Macintosh, and due to the small size of the network, and the relatively small wavelet transform, processed inputs very quickly. Behavior was also observed to be relatively smooth around the peak threshold and relatively stable. Overall performance of the neural net resulted in good generalization, with rates for testing showing only a small loss over training set accuracy.

Face	Number of Nodes			Constants			Percent Correct		Fixed
Algorithm	Input	Hidden	Output	Learning	Momentum	Sigmoid	Training	Test	Threshold
FACEIT	10	5	1	0.05	0.5	0.05	95.7	94.5	0.4
EBGM	10	5	1	0.05	0.5	0.5	95.2	92.4	0.45

Table 1. Backpropogation network architecture and performance.

4.3 Random Eye Perturbation Experiments

To study the effectiveness of our approach, we first analyze our prediction ability with respect to controlled simulation experiments. The images used in this experiment are from one session of outdoor data arbitrarily selected from our larger data set collected as follows. Each session consists of the same 1024 FERET images used for training, but displayed on an outdoor LCD monitor and re-acquired under varying time and weather conditions. Images are projected on a 15" LCD monitor and acquired asynchronously by two cameras at high speed from a distance of approximately 100 and 200 ft. Images are zoom adjusted so that facial images have approximately 50-100 pixels between the eyes. Eye coordinates for all images are computed, using the known location of the eyes from the gallery image and a pair of easily identifiable markers located in the projected image.

A series of random Gaussian offsets were applied to the eye coordinates of all images to create a series of probe sets with varying degrees of eye localization error. For this set of experiments, we selected offsets with a mean of zero and four different standard deviations: 2, 4, 6 and 8 pixels radially from the center of the known location of the eye. Note that different random perturbations were applied to each image, and 30 different random seeds were used for each standard deviation. This resulted in 4x30=120 runs of each algorithm on the same set of 1024 images. The intent of this experiment was to show the effectiveness of our approach as eye localization increases in error.

The data flow for the analysis of a single probe image is shown in figure 2. For each probe, the similarity scores are processed and the feature vector passed through the previously trained neural net. If neural net output exceeds the fixed threshold, the image is determined to have a high probability of being correctly classified and its classification is left intact. However, if the neural net output is below the threshold, the image is assumed to have a low probability of being correctly classified, and is then passed onto to the next stage of processing.

Three different subsets of eye offsets were investigated for their effectiveness. In a production setting, it may not be feasible to try all (for example) 81 combinations of offsets (or more) from a resource point of view. It would be beneficial to determine a smaller set of eye perturbations that have a high likelihood of achieving good performance gains versus the cost of reprocessing images. As a result, three subsets of eye offset combinations were tested, referred to as: SCALE (6 offsets), TRANSLATE_SCALE (26 offsets and X_SEP_CONSTANT (8 offsets).

SCALE included those offsets that simply increased or decreased the x separation between the eyes, embodying the implicit hypothesis that scaling is a significant factor affecting face algorithm performance.

X_SEP_CONSTANT included those offsets that simply translated the given x coordinates for both eyes, keeping the distance between them the same.

Finally, TRANSLATE_SCALE included all previous offsets, including scaling in conjunction with translation. No offsets in which one eye was translated in relation to the other were included in the analysis due to the prohibitive cost of post-processing.

Once a probe is identified as having a low probability of being correctly classified, it is then perturbed with an offset, and reprocessed by the face algorithm. This is repeated for all offsets in the subset. The feature vectors each time are input to the neural net, and the largest output (out of all of the offsets applied) is noted. The ranking information for this result supercedes the original classification only if:

1. its neural net output exceeds the fixed threshold
2. its neural net output exceeds that of the original

Results. First, it is instructive to look at how the algorithm behaves with respect to the decisions that are made during processing. As shown in figure 3, the neural net performs extremely well on the initial data, achieving a classification accuracy exceeding 90% over the entire range of initial input eye perturbation. Recall that the eye perturbation in this case is a random Gaussian variable and different for every single image, resulting in a rigorous test for the neural net. Note also, the very low false negative and false positive rates, indicating a relatively high efficiency (at least at this level) of the algorithm.

Not unexpectedly, as the variance of initial eye perturbation increases, performance decreases. However, it is interesting to note that there is a greater *relative* gain as variance increases, and as performance in general decreases. This is shown quite clearly in figure 4. This suggests that such a method might be even more useful as eye localization errors increase since at least one of the perturbations used to try to correct the classification error may be in the direction of the needed change. Changes in and around the correct location may not result in significant benefits. Nevertheless, even in the case of small initial perturbations, significant improvements (albeit small) were noted.

In general, TRANSLATE_SCALE performed slightly better than SCALE, but at a significantly higher cost (see figure 5). With only six offsets, SCALE was able to improve recognition performance significantly with much lower cost. This fact is not very surprising if one considers the importance of scaling in face analysis systems. These results suggest that adjusting factors that affect normalization(specifically eye separation distance) and then re-processing is a prudent approach to improving face recognition. This is consistent with observations made in [4] that eye separation

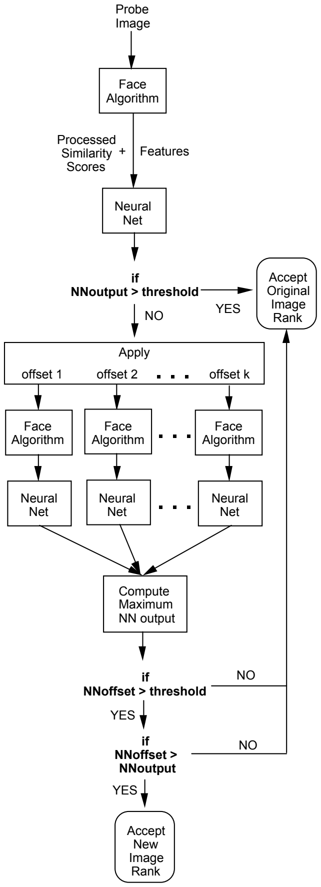

Figure 2. Flowchart showing data flow for the analysis of a single probe image

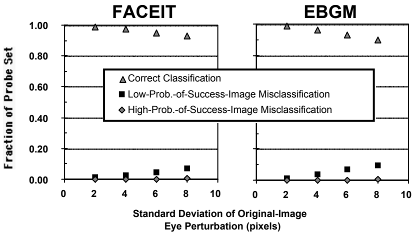

Figure 3. Classification accuracy of the neural network for FACEIT and EBGM algorithms

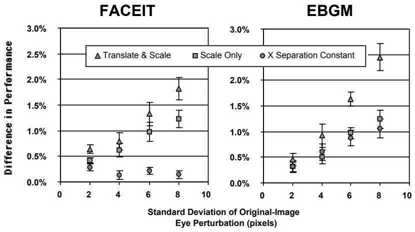

Figure 4. Percent difference in performance gain for various subsets of eye offsets

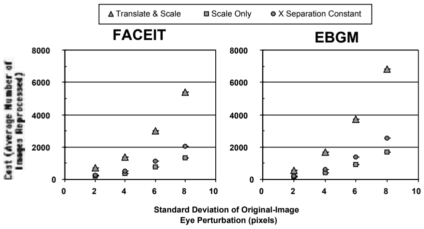

Figure 5. Maximum number of images reprocessed for each algorithm

distance seemed to have a greater effect on face recognition performance than the actual location of the eyes themselves.

Not surprisingly, X_SEP_CONSTANT performed considerably worse, although due to the accuracy of the neural net, performance did not degrade. It is conceivable that bad decisions by the neural net could result in falsely classifying an image as having a high probability of being classified correctly after applying an eye perturbation; however, this was clearly not the case.

With respect to the behavior of the neural network during processing, several important observations can be made. Results only for SCALE are shown in figure 6. First, the fraction of perturbed images that actually resulted in a degraded classification is extremely low, on the order of about 0.1%. Informal observations of the data indicated that even so, the amount of degradation was usually on the order of 1 or 2 ranks (e.g. changing a rank 1 image to a rank 2 or 3). Second, recall that once a probe is initially identified as having a high probability of being incorrectly classified, the image is offset multiple times and the output of the neural net for each re-processing is used to determine what to do with it. If the neural net determines the new result has a low probability of being correctly classified, that result is not considered. As seen in the top of figure 6, the fraction of perturbed images for which this is true is rather high. However, this is to be expected since the likelihood of a given perturbation to actually make things worse is rather high. In fact, the neural net is actually doing quite well, rejecting a large number and accepting only reasonably good possibilities. Of those accepted, *i.e.* when failure is predicted successfully (see the bottom of figure 6), approximately 50% result in an improvement in rank.

4.3 Biometric Perturbations of Real Images

Finally, a set of experiments shown in figure 7 clearly shows the benefit of the approach for real images. Four different times of day throughout the month of May were used for this analysis. SCALE perturbations were used to significantly improve face recognition results for the FACEIT algorithm. Note that in this set of experiments, errors in eye localization come from two sources: the eye localization error due to degradation of the input image as a result of atmospheric effects, and the eye localization error due to possible weaknesses in the FACEIT eye localization algorithm. Together, eye localization error is clearly an unknown quantity, but is exploited quite effectively here, to improve overall classification.

5 Conclusions

Eye localization has been shown to have a significant impact on face recognition algorithms. This paper uses that fact to show how machine learning and failure prediction can be integrated into a perturbation-based approach for overall system improvement. Our approach was tested on synthetic data using two different face-recognition systems; it showed both good failure prediction performance and, when failure was predicted, corrected for it about 50% of the time. It also managed to do so rather efficiently, requiring only a fraction of the total number of offset combinations, and would be expected to do even better in a production environment.

Using outdoor face data and a commercial face recognition system, the approach was able to predict failures and then predict which perturbations to keep, to achieve a

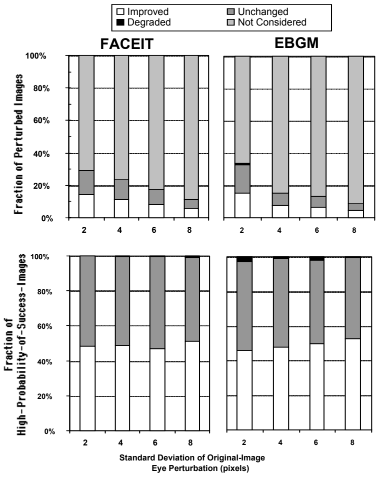

Figure 6. Breakdown of improved, degraded, unchanged and unconsidered (not detected as failures) images.

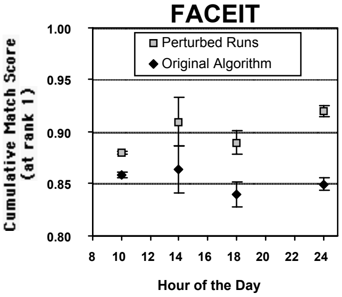

Figure 7. Performance of FACEIT before and after biometric perturbation. 95% confidence intervals are shown.

statistically significant 3% to 8% overall improvement beyond the already impressive 85% overall recognition rate of the base commercial face recognition system.

While this paper has focused on face recognition, since the use of "similarity measures" is ubiquitous, this approach should apply across a broad range of pattern recognition problems. In fact, any instance where a weak link exists in a pattern recognition problem, and that also has a limited local perturbation space, is a viable candidate for such an approach.

References

1. Brunelli, R. and Poggio, T. (1993). Face Recognition: Features versus Templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, pp. 1042-1052.
2. Valentin, D., Abdi, D., O'Toole, J., and Cottrell, G. (1994). Connectionist Models of Face Processing: A Survey. *Pattern Recog.*, **27**, pp. 1209-1230.
3. Riopka, T.P. and Boult, T. (2003). The Eyes Have It. *Proceedings of the ACM Biometrics Methods and Applications Workshop*, Berkeley, CA. pp. 33-40.
4. Marques, J., Orlans, N.M., and Piszcz, A.T. (2003). Effects of Eye Position on Eigenface-Based Face Recognition Scoring. *Technical Paper*, Mitre Corp.
5. Okada, K., Steffens, J., Maurer, T., Hong, H., Neven, H. and von der Malsburg, C. (1998). The Bochum/USC Face Recognition System and How It Fared in the FERET Phase III Test. In *Wechsler et al.*, eidtores*, Face Recognition: From Theory to Applic.*, pp. 186-205.
6. Penev ,P. S. and Atick, J. J. (1996). Local feature analysis: A general statistical theory for object representation. *Neural Systems*, 7:477-500.
7. Bolme, D.S., Beveridge, J.R., Teixeira, M. and Draper, B.A. (2003). The CSU Face Identification Eval. System: Its Purpose, Features, and Structure. *ICVS 2003*: 304-313.
8. Daubechies, I. (1988). Orthonormal Bases of Compactly Supported Wavelets. *Comm. Pure Appl. Math.*, 41, pp. 909-996.
9. McClelland, J. and Rumelhart, D. (1986). Explorations in Parallel Distributed Processing, Volumes 1 and 2. MIT Press, Cambridge, MA.