Fusion of Multiple Facial Regions for Expression-Invariant Face Recognition

Wei-Yang Lin, Ming-Yang Chen
Department of Computer Science and Information Engineering
National Chung Cheng University
Min-Hsiung, Chia-Yi, Taiwan
Email: wylin@cs.ccu.edu.tw

Kerry R. Widder, Yu Hen Hu, Nigel Boston
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
Madison WI, 53706 USA
Email: {widder, hu, boston}@wisc.edu

***Abstract*— In this paper, we describe a fusion-based face recognition method that is able to compensate for facial expressions even when training samples contain only neutral expression. The similarity metric between two facial images are calculated by combining the similarity scores of the corresponding facial regions, *e.g.* the similarity between two mouthes, the similarity between two noses, etc.. In contrast with other approaches where equal weights are assigned on each region, a novel fusion method based on linear discriminant analysis (LDA) is developed to maximize the verification performance. We also conduct a comparative study on various face recognition schemes, including the FRGC baseline algorithm, the fusion of multiple regions by sum rule, and the fusion of multiple regions by LDA. Experiments on the FRGC (Face Recognition Grand Challenge) V2.0 dataset, containing 4007 face images recorded from 266 subjects, show that the proposed method significantly improves the verification performance in the presence of facial expressions.**

I. Introduction

The expression variations is an inevitable issue in the development of a practical face recognition system. However, most of the researches to date deal with other difficulties, such as illumination or pose changes [1]. The issue can be described as follows:

> Given facial images with different expressions, how could we devise an algorithm which robustly identify a person's face?

In this paper, we propose a multi-region approach for improving the robustness to facial expressions. Intuitively, we argue that smaller facial regions, if judiciously selected, would be less sensitive to expression variations and may lead to better overall performance. A key research issue in a multi-region approach is to devise an effective fusion method so that individual classification results based on different facial regions can be combined to yield the best result. Along this direction, we proposed a score-level information fusion approach: a weighted combination of similarity scores based on Fisher's Linear Discriminant Analysis (LDA) [2]. The experimental validation of the proposed approach is carried out by using the Face Recognition Grand Challenge (FRGC) version 2.0 dataset [3] and the Biometric Experimentation Environment (BEE) accompanying FRGC.

The rest of this paper is organized as follows: In section II, we will briefly introduce the Face Recognition Grand Challenge [3] and give details of our face recognition method which utilizes multiple facial regions. In section III, we present the empirical results of various methods, including the FRGC baseline, the single-region methods, and the multi-region methods. Section IV gives a discussion of issues raised by the approach presented in this paper. Finally, we make concluding remarks in section V.

II. The Dataset and the Proposed Method

A. The FRGC v2.0 dataset

We perform face recognition experiments on the FRGC v2.0 dataset [3]. In the FRGC dataset, 3D images with a resolution 640×480 consist of both shape and textures channels. The two disjoint data partitions, training and validation partitions, contains 943 and 4007 3D images respectively. Among the challenge problems defined in FRGC v2.0, we focus on experiment 3t which utilizes only texture channel in a 3D image. Performance will be reported on a Receiver Operator Characteristic (ROC) that shows the trade-off between verification rate and false accept rate (FAR). In the FRGC protocol, three masks are defined over the similarity matrix where each entry contains the matching score of an image pair. Each mask collects its own set of entries in a similarity matrix, thus generating three ROC curves which will be referred to as ROC I, II and III. The images within an image pair are called gallery and probe. In ROC I, gallery and probe are from the same semester. In ROC II, gallery and probe are from the same year. In ROC III, gallery and probe are from different semesters. In average, ROC III has the longest time lapse between gallery and probe and therefore is the most challenge experiment. Throughout this paper, we use ROC III to compare the performances of different methods, unless otherwise specified.

B. Proposed Face Recognition Method

In order to identify the improvement achieved by using multiple facial regions, the proposed face recognition method closely follows the structure of the FRGC baseline algorithm. Interesting readers may refer to [4] for details about the FRGC baseline algorithm. The step-by-step illustration of our method is as follows:

1) *Texture only experiment*: We intentionally discard the shape images in our experiments. Hence, the resulting performance is completely relied on the features extracted from texture images.
2) *Preprocessing*: The main objective of the preprocessing is to normalize the geometry and the brightness of a facial image. A 480×640 color image is converted to a 150×130 graylevel image during this step, as shown in Figure 1. The geometric normalization gives the same pixel distance between eye locations to all faces. Also, the nose tip of every subject is translated to the same location on a normalized image. To compensate for brightness changes, we apply the histogram equalization to reduce the effect of illumination variations. Note that we use the implementation provided by the FRGC baseline algorithm without any modification.
3) *Specify the regions of interest*: Instead of using the entire normalized texture image as the FRGC baseline algorithm does, the feature vectors are extracted from local regions. We use 3 regions in our experiments, namely forehead, nose, and left eye (see Figure 2). These areas tend to change less for expression variations.
4) *Dimension reduction*: We use a popular technique, the Principal Component Analysis (PCA) [5], [6], to obtain a linear subspace for each region. The basis of linear subspaces is computed by using the training partition containing 943 range images. The first 10 eigenvectors are dropped and the following 100 eigenvectors span a lower dimensional feature space. Note that the parameters for dimension reduction are exactly the same as those used in the FRGC baseline algorithm.
5) *Similarity metric* : Given two feature vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$, the similarity between these two feature vectors, denoted by $s(\mathbf{a}, \mathbf{b})$, is defined as the following.

$$s(\mathbf{a}, \mathbf{b}) = \frac{\sum_{i=1}^{d} \frac{1}{\sqrt{\lambda_i}} a_i b_i}{\|\mathbf{a}\|_M \cdot \|\mathbf{b}\|_M} \tag{1}$$

where

$$\|\mathbf{a}\|_M^2 = \sum_{i=1}^{d} \frac{1}{\sqrt{\lambda_i}} a_i^2 \tag{2}$$

The λ_i denotes the i^{th} largest eigenvalue of the covariance matrix [7]. This is also the similarity metric used by the FRGC baseline.
6) *Fusion of similarity metrics*: There are many methods for conducting fusion on biometrics [4], [8], [9], which usually lead to a better identification/verification performance. Here, we use two simple schemes, namely the *sum rule* and the *linear discriminant analysis*, to combine the similarity metrics from different facial regions. The hypothesis is that combining multiple regions will deliver a consensus decision with better quality.

Fig. 1. The scale and position of a face are normalized during the preprocessing step. *Left*: a raw image; *Right*: the resulting normalized image.

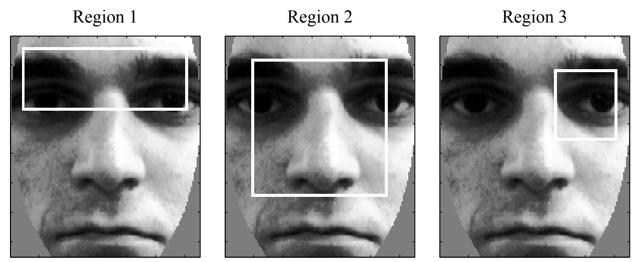


Fig. 2. We specify 3 regions on a facial image. Matching scores obtained from each region are combined to yield the final matching score.

III. Experimental Results

A. The Single Region Algorithms

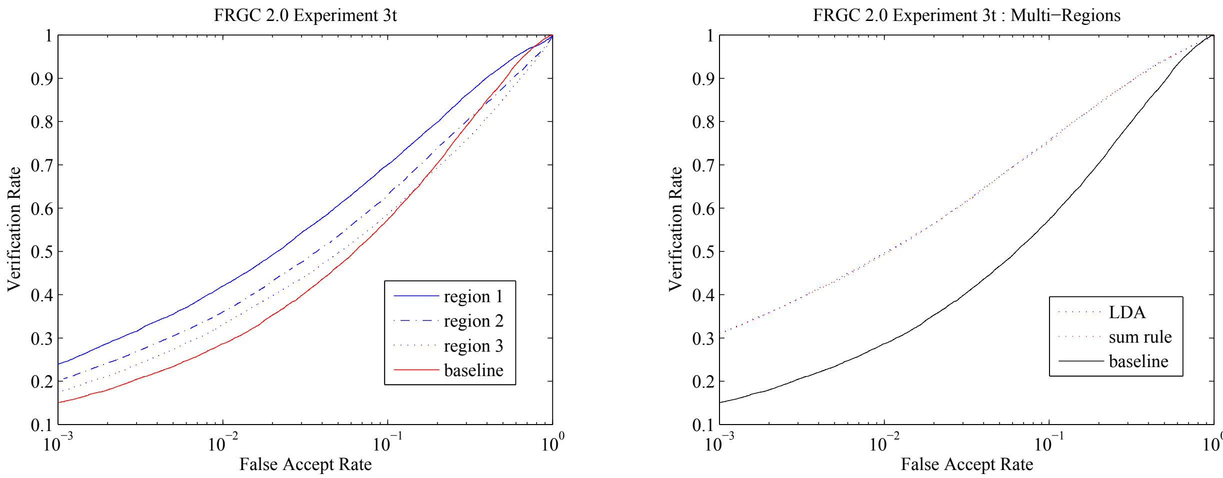
Before combining facial regions, we first conduct face recognition experiments by using individual facial region. The ROC curves of three different regions are shown in Figure 3(a). The verification rates at 0.1% FAR are 23.96%, 20.22%, and 17.60% respectively for forehead, nose, and left eye regions. All of the selected regions yield better results than that of the FRGC baseline. The verification rate at 0.1% FAR of the FRGC baseline is 15.03%. Obviously, a relatively small facial region is more robust to expression variations than a whole face. Hence, one can improve the face recognition performance by using a local facial region.

Intuitively, some facial regions should be more important than others in terms of face recognition performance. To explore the relative importance of each region, we could compare their ROC curves in Figure 3(a). The results suggest that forehead region yields the highest verification rate followed nose region. This is an interesting result since psychological experiments typically indicate that eyes are the most important followed by mouth and the nose [10].

So, how could we reasonably interpret this empirical finding? One plausible explanation is that the shape of mouth and eyes will change across different expressions. If a face recognition algorithm utilize these regions, the intra-class variation would increase in the presence of expression variations.

B. Fusion by Sum Rule

The simplest way of fusion is to add the similarity scores from different regions together, *i.e.* equal weights are assigned since we don't know the relative importance of each region.



(a)

(b)

Fig. 3. ROC curves obtained by (a) using different facial regions and (b) combining multiple facial regions.

Figure 3(b) shows the results of sum rule on the FRGC experiment 3t. Note that the verification performance of sum rule is better than that of any single region. This confirms the fact that a better classification performance can be achieved by combining multiple classifiers. For ROC III, the verification rate of sum rule is 30.84% at 0.1% FAR, which is a significant improvement over the FRGC baseline.

C. *Fusion by Linear Discriminant Analysis*

In the FRGC protocol, there are two non-overlapping data partitions: training and validation. During algorithm development, experiments are conducted on the validation partition. This allows researchers to tune the parameters of their approaches. However, finding a subspace representation and classifier training are required to be conducted on the training partition. Here, we use Fisher's Linear Discriminant Analysis (LDA) [2] to obtain the optimal weighting vector from the training samples. In FRGC v2.0 dataset, there are totally 943 images in the training partition.

Let $\mathbf{s} = (s_1, s_2, \ldots, s_n)^T$ denote a score vector, where n is the number of facial regions, $n \leq 3$ in our experiments. Each s_i is the matching score of i^{th} region, generated from a pair of images. The goal of discriminant analysis is to find the optimal projection direction for which the projected samples are well separated. A projected sample π is a linear combination of the components of $\mathbf{s}$.

$$\pi = \mathbf{w}^T \mathbf{s} \tag{3}$$

where $\mathbf{w}$ is the weighting vector that maximizes the following criterion function.

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \tag{4}$$

The *between-class scatter matrix* $\mathbf{S}_B$ and the *within-class scatter matrix* $\mathbf{S}_W$ are defined by

$$\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \tag{5}$$

$$\mathbf{S}_W = \sum_{i=1}^{2} \sum_{\mathbf{s}_k \in \mathbb{S}_i} (\mathbf{s}_k - \mathbf{m}_i)(\mathbf{s}_k - \mathbf{m}_i)^T \tag{6}$$

Here, we have a typical two-categories classification problem. The two classes are the *matching* class and the *non-matching* class. The matching class contains the score vectors from image pairs of the same person while the non-matching class is generated by image pairs of different persons. In Eq. (5) and (6), $\mathbf{m}_1$ is the mean score vector of training samples belonging to match class and $\mathbf{m}_2$ is the mean score vector of training samples belonging to non-match class. Similarly, $\mathbb{S}_1$ denotes the set of training samples belonging to match class and $\mathbb{S}_2$ denotes the set of training samples belonging to non-match class. The $\mathbf{w}$ that maximize the criterion function $J(\mathbf{w})$ is given by

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \tag{7}$$

The LDA provides the optimal projection direction $\mathbf{w}$ that maximizes the separation between projected samples. The result of fusion by LDA is shown in Figure 3(b). Once again, we observe a significant improvement over the FRGC baseline. We should mention the fact that the training partition contains only neutral expression while the validation partition contains different expressions. Hence, the distribution of training samples is expected to be different from the distribution of validation data in the FRGC v2.0 protocol. This points out a future direction that one could expect performance improvement by including facial images with expression changes in the training partition. The weights assigned on the selected regions are shown in Table I.

TABLE I

THE WEIGHTS ON THE SELECTED REGIONS OBTAINED BY LDA

Number of selected regions	Weights			VR @ FAR = 0.1%		
	#1	#2	#3	ROC-I	ROC-II	ROC-III
3 regions	0.5364	0.6918	0.4834	0.4426	0.3817	0.3104
2 regions	0.5928	0.8054	0	0.4134	0.3529	0.2816
1 region	1	0	0	0.3599	0.3045	0.2396
1 region	0	1	0	0.3082	0.2596	0.2022
1 region	0	0	1	0.2566	0.2195	0.1760
FRGC baseline algorithm				0.2793	0.2204	0.1503

IV. DISCUSSION

From the experiment results in Figure 3(b), we observe that the LDA achieves similar performance as the sum rule. For ROC III, the verification rate of LDA and sum rule are 31.04% and 30.84% respectively (at FAR = 0.1%). Ideally, the LDA obtains the optimal weighting vector from the training samples. It is expected that the distribution of training data is an important prior knowledge for us to construct a better fusion scheme. In the FRGC experiment 3t, however, the representativeness of training samples is quite limited since there is only neutral expression in the training images. Even with a vast amount of training data, the resulting weighting vector would be unlikely to do well on novel images.

Rather, we might seek to simplify our fusion strategy, motivated by a belief that the similarity scores from different regions are independent and identically distributed (i.i.d.). Under the i.i.d. assumption, it is easy to show that the weighting vectors obtained by the LDA and the sum rule are equivalent. Indeed, a simpler fusion method might have slightly poorer performance on the training samples. But, if a sophisticated fusion scheme is unlikely to yield good generalization, we would favor the simplest combination rule, *i.e.* the sum rule. Similar observations have been reported by Kittler *et al.* [11]. They conducted a comparative study on various classifier combination schemes and found that the sum rule outperformed others.

V. CONCLUSION

We have developed a face recognition system which integrates multiple regions of a facial image. The proposed system overcomes the limitations of the single-region algorithm. The performance improvement is due to the combination schemes which generate the final matching score with a higher quality than those based on a single region. Experimental results indicate that the verification performance of the fusion by LDA is the same as the fusion by sum rule. One possible explanation is that the distribution of training samples is different from the distribution of testing samples in the FRGC v2.0 dataset. This is the issue of *generalization* [2]. If the training samples are not representative, their availability is not helpful for us to build a better fusion scheme. Hence, it is unlikely that a complex fusion scheme with good training accuracy would also provide good generalization. Overall speaking, the proposed multi-region face recognition algorithm yields promising results on the FRGC v2.0 dataset, especially the improved robustness in the presence of facial expressions.

ACKNOWLEDGMENT

This work has been supported by US National Science Foundation under the grant No. CCF-0434355 and the National Science Council, Taiwan, under the grant No. 95-2218-E-194-015.

REFERENCES

[1] W. ZHAO, R. CHELLAPPA, P. PHILLIPS, and A. ROSENFELD, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[2] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Wiley-Interscience, 2000.

[3] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 947–54.

[4] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D+3D face biometrics," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619–24, 2005.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neurosccience*, vol. 3, no. 1, pp. 71–86, 1991.

[6] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for thecharacterization of human faces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 1, pp. 103–108, 1990.

[7] W. Yambor, "Analysis of PCA-based and Fisher discriminant-based image recognition algorithms," Master's thesis, Colorado State University, 2000.

[8] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003.

[9] S. Kung and M. Mak, "On Consistent Fusion of Multimodal Biometrics," in *Acoustics, Speech, and Signal Processing, 2006. Proceedings.(ICASSP'06). IEEE International Conference on*, vol. 5, 2006, pp. 1085–1088.

[10] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: 19 results all computer vision researchers should know about," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006.

[11] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 3, pp. 226–239, 1998.