3D FACE RECOGNITION UNDER EXPRESSION VARIATIONS USING SIMILARITY METRICS FUSION

Wei-Yang Lin

National Chung Cheng University
Department of CSIE
Min-Hsiung, Chia-Yi, Taiwan

*Kin-Chung Wong, Nigel Boston, Yu Hen Hu**

University of Wisconsin-Madison
Department of ECE
Madison WI, 53706 USA

ABSTRACT

We present a novel 3D face recognition method that incorporates summation invariant features extracted from multiple sub-regions of a facial range images, and optimal fusion of similarity scores between corresponding sub-regions. The key innovation of this paper is the development of the fusion-based face recognition algorithm that delivers significant performance enhancement while requiring very little computation. Experiments on the FRGC (Face Recognition Grand Challenge) version 2 dataset show that our algorithm improves the recognition performance significantly in the presence of facial expressions.

1. INTRODUCTION

Face recognition based on 3D information has received unprecedented interest in recent years [1]. However, few results have been reported on dealing with the variations of facial expressions. Like variations caused by illumination and pose changes, expression variation is another fundamental issue in face recognition. A practical face recognition solution has to deal with a certain degree of expression changes even with cooperative users. The advantages of 3D facial features include that illumination variations may be of lesser problem and pose variations can be normalized provided facial landmarks are given [2]. Therefore, facial expressions eventually become the bottleneck to current 3D face recognition systems.

From the comprehensive survey on 3D face recognition by Bowyer *et al.* [1], there are relatively fewer research works reported on handling the great challenge posed by facial expressions. Passalis *et al.* [3] propose a fully automatic 3D face recognition algorithm based on the elastically adapted deformable model frame work. Their deformable face model can perform non-rigid alignment with input facial surface under different expressions. A 3D face recognition approach based on the invariant feature of isometries was introduced by Bronstein *et al.* [4]. The key innovation of their approach lies in utilizing geodesic distance between the fiducial points on a 3D face surface. Chang *et al.* [5] tackle the issue of expression changes by utilizing three different nose regions of a face, as being relatively rigid areas across different facial expressions.

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 pattern 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: an optimal linear weighted fusion approach based on classical Linear Discriminant Analysis (LDA). The experimental validation of the proposed approach is carried out by using the Face Recognition Grand Challenge (FRGC) version 2 dataset [6] and the Biometrics Experimentation Environment (BEE) accompanying FRGC.

The remainder of this paper is organized as follows: In section 2, we will briefly introduce the single region 3D face recognition method developed in [7]. The single region algorithm serves as the building block of the proposed multi-region 3D face recognition algorithm. In section 3, we present two fusion schemes, which combine all the facial sub-regions and yield final results, and the empirical results. Finally, concluding remarks are made in section 4.

2. SINGLE-REGION 3D FACE RECOGNITION

In the UW-Madison face recognition laboratory, we have developed novel families of geometrically invariant features, known as summation invariants [8, 9], and apply them for 3D face recognition tasks. In this section, we briefly review the single region 3D face recognition approach using summation invariants.

Given a 3D facial range image and the locations of fiducial points on the face surface, we can identify an 81×81 rectangular region centered at the nose tip as the region of in-

*The authors have been partially supported by US National Science Foundation (Grant No. CCF-0434355) and the National Science Council, Taiwan (Grant No. 95-2218-E-194-015).

terest. Note that FRGC provides the locations of eye corners, nose tip and mouth corners. A normalized range image and the 81×81 region used in our single region method are shown in Figure 1. Each row and each column of this rectangular region of the range image is a 2D curve. In this work, we only compute the summation invariant $\eta_{1,1}$, which is a member of the Euclidean summation invariant family, from a 2D curve. Given N points over a curve $\{(x_n, y_n)\}$, we have [7]:

$$\begin{aligned}
\eta_{1,1} &= P_{1,1}((x_1 - x_N)^2 - (y_1 - y_N)^2) + P_{1,0}(y_1^3 \\
&+ 2x_1 x_N y_N - 2y_N x_1^2 + x_1^2 y_1 - 2y_1^2 y_N + y_1 y_N^2 \\
&- x_N^2 y_1) - P_{0,1}(x_1^3 + 2y_1 y_N x_N - 2x_N y_1^2 + y_1^2 x_1 \\
&- 2x_1^2 x_N + x_1 x_N^2 - y_N^2 x_1) \\
&+ (P_{0,2} - P_{2,0})(x_1 - x_N)(y_1 - y_N) \\
&+ N(x_N y_1 - x_1 y_N)(x_1(x_N - x_1) + y_1(y_N - y_1))
\end{aligned}$$

where the *potentials* are given by $P_{i,j} = \sum_{n=1}^{N} x_n^i y_n^j$. Instead of using the entire 2D curve to compute a single summation invariant, we compute $\eta_{1,1}$ over a local segment surrounding each point on a 2D curve . The length of the local segment is chosen to be 21 in our experiments. This semi-local summation invariant for a 2D curve, thus yields another 2D curve, with fewer points due to the local segment. Combining all 2D feature curves, the single region of range image yields a high-dimensional feature vector, called summation invariant feature vector. We then apply the Principal Component Analysis (PCA) to reduce the dimensionality of the summation invariant feature vector. After proper dimension reduction, we obtain a compact representation for the 81×81 region of a given facial range image. Interested readers may refer to [7] for details on the feature extraction of 3D range images.

As a participating group of the Face Recognition Grand Challenge (FRGC) [6], we follow the designated experiment protocol to compute a similarity score for each pair of images, one in the gallery set and the other in the probe set. If the image pair belong to the same person, the similarity score is labeled as a *match* score. Otherwise, it is labeled as a *non-match* score. After all images in the gallery and in the probe set are compared, the statistical distributions of the match scores and the non-match scores can be derived. These distributions then give a Receiver-Operating Characteristic (ROC) curve which is used to represent the performance of a particular algorithm.

In the FRGC protocol, three masks are defined over the similarity matrix where each entry contains the similarity score of an image pair. Each mask collects its own set of entries in the similarity matrix, thus generating three ROC curves which will be referred to as ROC I, II and III. In ROC I, the gallery image and probe image 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 challenging experiment. More details of FRGC experiments can be found in [6].

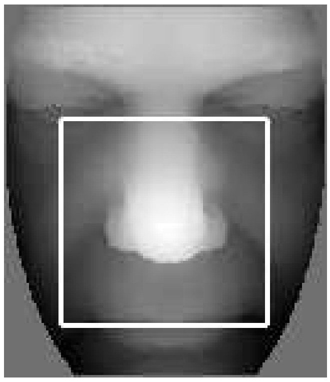


Fig. 1. A normalized range image and the 81×81 region centered at nose tip.

The single region algorithm works very well on the FRGC v1.0 dataset which contains only neutral expression. The resulting verification rate is 97.2%, measured at 0.1% false accept rate [7]. Unfortunately, we observe a significant performance drop when it was applied to the FRGC v2.0 dataset where expression variations exist. The results of applying single region algorithm on the FRGC v2.0 dataset are shown in Figure 3(a). At 0.1% false accept rate, the verification rates for ROC I, II and III are about 75%, 73% and 71.5%, respectively. In this work, we attempt to address this issue with a multi-region approach.

3. MULTI-REGION 3D FACE RECOGNITION

Our strategy for multi-region 3D face recognition can be highlighted as follows: We will partition a 3D facial range image into 10 rectangular sub-regions, as shown in Figure 2. These regions are located on the normalized range images, which are provided by the FRGC baseline algorithm. Note that the location of nose tip and the pixel distance between eye corners are normalized during the preprocessing stage [2]. Hence, one can then specify the size and location of a selected region by using the pixel coordinate frame. For each sub-region, a single-region face recognition procedure as described in the previous section will be performed. This yields a similarity score that specifies how similar this given sub-region of a gallery range image is to the same sub-region of a probe range image. We then combine the similarity scores of all sub-regions to yield a fused similarity score which then can be used to deduce the performance of the proposed method. The main innovation of this research is the development of novel score-level information fusion approaches.

3.1. Score-level Information Fusion

For each pair of 3D range images, 10 similarity scores are computed from the 10 selected sub-regions. A 10×1 score

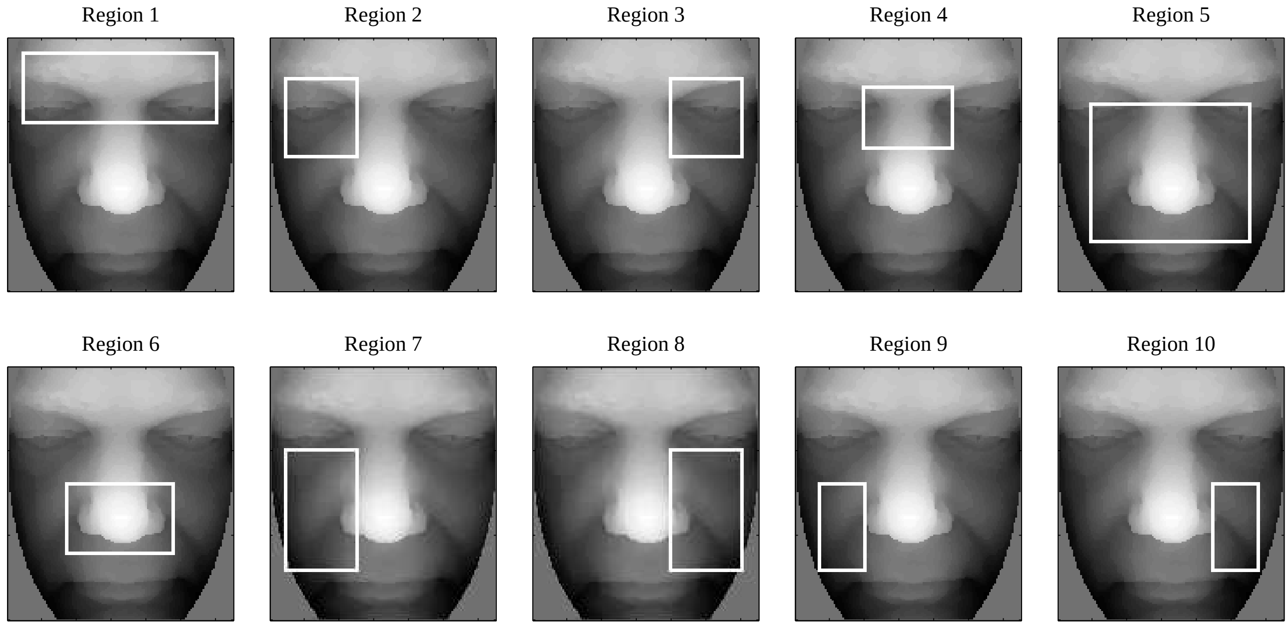


Fig. 2. We specify 10 regions on a facial surface. Matching scores obtained from each region are combined to yield the final matching score.

vector is obtained by simply stacking 10 similarity scores together. Each of these score vectors can be regarded as a derived feature vector that has a label of either *match*, if an image pair are from the same person, or *non-match*, if an image pair are from different persons. In essence, this is a typical two-category classification problem. The objective of fusion is to devise a function from $\mathbb{R}^{10}$ to $\mathbb{R}$ such that those labeled with *match* will be mapped to a positive value and those labeled with *non-match* will be mapped to a negative value. In this paper, we use linear fusion methods to achieve this goal. More specifically, one can compute the fused score by using a weighted sum of individual similarity scores in a score vector. In the following, we will present two fusion schemes and their performance on FRGC experiment 3s. Note that among challenge problems defined in FRGC, we focus on experiment 3s which utilizes only facial range images.

3.2. Sum Rule

The simplest way of fusion is to add the similarity scores from all the selected sub-regions. In other words, equal weights are assigned on each sub-region since we don't know the relative importance of these sub-regions. Figure 3(a) shows the results of sum rule on the FRGC v2.0 dataset. By combining relatively small regions which are relatively unaffected by expressions, we observe a significant improvement over the single region method in terms of verification rate. At a 0.1% false accept rate, the verification rates for ROC I, II and III are about 90%, 89% and 88%, respectively.

3.3. Optimal Linear Fusion Using LDA

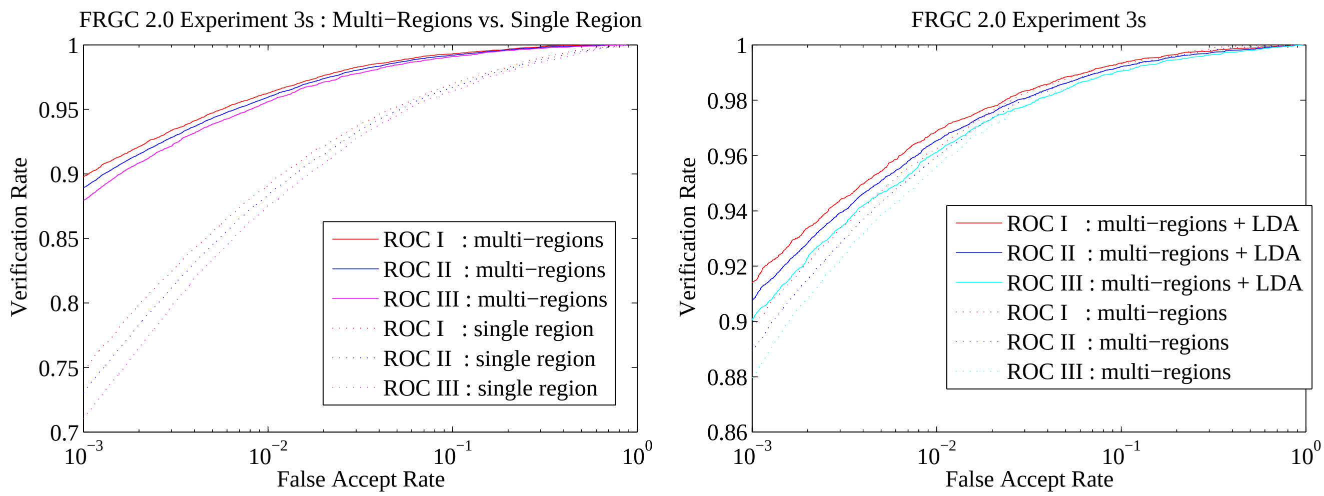
The classical Linear Discriminant Analysis [10] attempts to find the optimal linear weight W_{LDA} that maximizes a Rayleigh quotient:

$$W_{LDA} = \arg\max_{W} \frac{W^T S_B W}{W^T S_W W}$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix. It is well-known that the solution has the form

$$W_{LDA} = S_W^{-1}(m^+ - m^-)$$

where m^+ is the mean score vector of training samples belonging to *match* class and m^- is the mean score vector of training samples belonging to *non-match* class. The FRGC protocol specifies two non-overlapping data partitions: training and validation. In FRGC v2.0 dataset, there are total 943 3D images in the training partition. Here, we use samples in the training partition to obtain the optimal weighting vector W_{LDA}. Figure 3(b) shows the ROC curves of using sum rule and LDA. Compared with fusion by sum rule, fusion by LDA provides an uniform improvement in verification rate at any false accept rate. In particular, at 0.1% false accept rate, fusion by LDA can achieve the verification rates of 91.5%, 91% and 90% for ROC I, II and III, respectively.



(a) (b)

Fig. 3. ROC performance obtained by information fusion. (a) Fusion by sum rule (solid lines) and single region method (dashed lines). (b) Fusion by LDA (solid lines) and sum rule (dashed lines).

4. CONCLUSION

We have developed a 3D face recognition system which integrates multiple sub-regions on a facial surface. The fusion of multiple sub-regions can provide a significantly better face identification result under a variety of expressions. In addition, the LDA is adopted to obtain the optimal weights which are assigned on each sub-region to maximize the verification rate. The proposed multi-region algorithm overcomes the limitations of the previous single-region algorithm which suffers performance degradation in the presence of expression variations. Experiment results on FRGC v2.0 dataset demonstrate that our algorithm is robust to facial expressions. The performance improvement is due to the fact that the information fusion scheme generates a final decision with higher quality than the decision based on a single classifier.

5. REFERENCES

[1] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1 – 15, 2006.

[2] 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, April 2005.

[3] I. A. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza, "Multimodal face recognition: Combination of geometry with physiological information," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 1022–1029.

[4] A.M. Bronstein, M.M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *Int. J. Comput. Vision*, vol. 64, no. 1, pp. 5–30, 2005.

[5] K.I. Chang, K.W. Bowyer, and P.J. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1695–1700, 2006.

[6] 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*, 2005, vol. 1, pp. 947–54.

[7] W. Y. Lin, K. C. Wong, N. Boston, and Y. H. Hu, "Fusion of summation invariants in 3D human face recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2006, vol. 2, pp. 1369 – 1376.

[8] W. Y. Lin, N. Boston, and Y. H. Hu, "Summation invariant and its application to shape recognition," in *Proc. IEEE Conf. on ICASSP*, 2005, vol. V, pp. 205–208.

[9] W. Y. Lin, K. C. Wong, N. Boston, and Y. H. Hu, "3D human face recognition using summation invariants," in *Proc. IEEE Conf. on ICASSP*, 2006, vol. 2, pp. 341 – 344.

[10] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, 2 edition.