AN INTEGRATED 3D FACE-EXPRESSION RECOGNITION APPROACH

Chao Li, Armando Barreto
Electrical & Computer Engineering Department
Florida International University

ABSTRACT

Face recognition has been a focus in research for the last couple of decades because of its wide potential applications and its importance to meet the security needs of today's world. With the development of 3D imaging technology, 3D face recognition emerges as an alternative to overcome the difficulties inherent to 2D face recognition, i.e. sensitivity to illumination conditions and positions of a subject. But 3D face recognition still needs to tackle the problem of deformation of facial geometry that results from the expression changes of a subject. To deal with this issue, a 3D face recognition framework is proposed in this paper. It is composed of three subsystems: expression recognition system, expressional face recognition system and neutral face recognition system. A system for the recognition of faces with one type of expression (smile) and neutral faces was implemented and tested on a database of 30 subjects. The results proved the feasibility of this framework.

1. INTRODUCTION

Most of the face recognition attempts that have been made until recently use 2D intensity images as the data format for processing. Detailed and comprehensive surveys can be found in [1, 2]. In spite of the success reached by 2D recognition methods, certain problems still exist, since 2D face images not only depend on the face of a subject, but also depend on imaging factors, such as the environmental illumination and the orientation of the subject. These variable factors can make the 2D face recognition system fail.

With the development of 3D imaging technology, more and more attention has been directed to 3D face recognition, which is robust with respect to illumination variation and posing orientation. In [3], Bowyer et al. provide a survey of 3D face recognition technology. Most of the 3D face recognition systems treat the 3D face surface as a rigid surface. But actually, the face surface is deformed by different expressions of the subject, which can cause systems that treat the face as a rigid surface to fail. The involvement of facial expression has become a big challenge in 3D face recognition systems. In this paper, we propose an approach to tackle this problem, through the integration of expression recognition and face recognition in a system.

2. EXPRESSION AND FACE RECOGNITION

From the psychological point of view, it is still not known whether facial expression recognition information aids the recognition of faces by human beings. In [4], Etcoff and Magee found that people are slower in identifying happy and angry faces than they are in identifying faces with neutral expression.

The proposed framework involves an initial assessment of the expression of an unknown face, and uses that assessment to facilitate its recognition. The incoming 3D range image is processed by an expression recognition system to find the most appropriate expression label for it. The expression labels include the six prototypical expressions of the faces, which are happiness, sadness, anger, fear, surprise and disgust[5], plus the neutral expression. According to different expressions, a matching face recognition system is then applied. If the expression is recognized as neutral, then the incoming 3D range image is directly passed to the neutral expression face recognition system, which uses the features of the probe image to directly match those of the gallery images, which are all neutral, to get the closest match. If the expression found is not neutral, then for each of the six prototypical expressions, a separate face recognition subsystem should be used. The system will find the right face through modeling the variations of the face features between the neutral face and the face with expression. Figure 1 shows a simplified version of this framework. This simplified diagram only deals with the smiling expression, which is the most commonly displayed by people publicly.

3. DATA ACQUISITION AND PREPROCESSING

To test the idea proposed in this model, a database, which includes 30 subjects, was built. In this database, we test the different processing of the two most common expressions, i.e., smiling versus neutral. Each subject participated in two sessions of the data acquisition process, which took place in two different days. In each session, two 3D scans were acquired with a Polhemus Fastscan scanner [6]. One was a neutral expression; the other was a happy (smiling) expression. The resulting database contains 60 3D neutral scans and 60 3D smiling scans of 30 subjects.

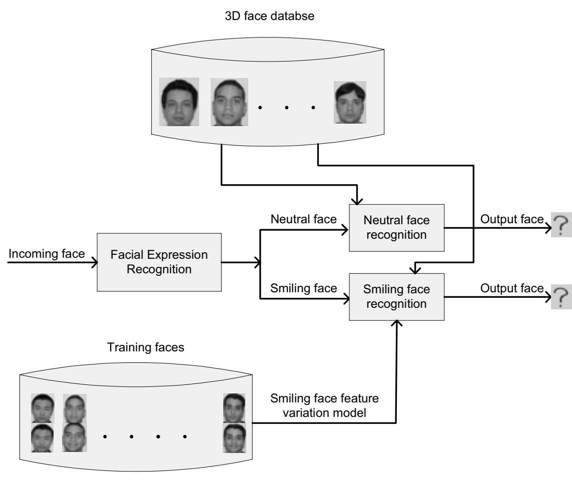

Figure1 Simplified framework of 3D face recognition

The left image in Figure 2 shows an example of the 3D scans obtained using this scanner, the right image is the 2.5D range image used in the algorithm, which was obtained by preprocessing as described in [8].

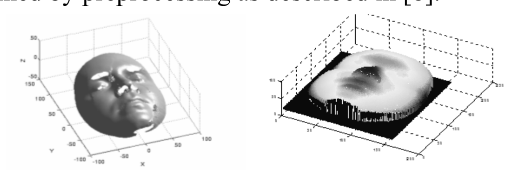

Figure 2 3D surface (left) and a mesh plot of the converted range image (right)

4. EXPRESSION RECOGNITION

The expression of the face is a basic mode of nonverbal communication among people. In [5], Ekman and Friesen proposed six primary emotions. Each possesses a distinctive content together with a unique facial expression. These six emotions are happiness, sadness, fear, disgust, surprise and anger. Together with the neutral expression, they also form the seven basic prototypical facial expressions.

In our experiment, we aim to recognize social smiles, which were posed by each subject. Smiling is generated by contraction of the zygomatic major muscle. This muscle lifts the corner of the mouth obliquely upwards and laterally, producing a characteristic "smiling expression". So, the most distinctive features associated with the smile are the bulging of the cheek muscle and the uplift of the corner of the mouth, as shown in Figure 3.

The following steps are followed to extract six representative features for the smiling expression:

1. An algorithm is developed to obtain the coordinates of five characteristic points in the face range image as shown in Figure 3. A and D are the extreme points of the base of the nose. B and E are the points defined by the corners of the mouth. C is in the middle of the lower lip.

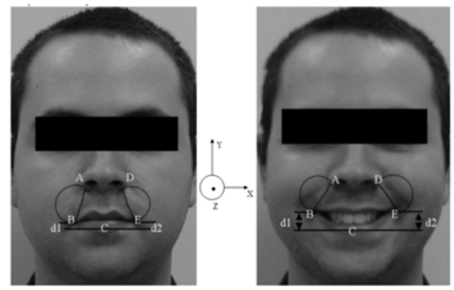

Figure 3 Illustration of features of a smiling face versus a neutral face

2. The first feature is the width of the mouth, BE, normalized by the length of AD. Obviously, while smiling the mouth becomes wider. The first feature is represented by *mw*.
3. The second feature is the depth of the mouth (The difference between the Z coordinates of point B point C and point E point C) normalized by the height of the nose to capture the fact that the smiling expression pulls back the mouth. This second feature is represented by *md.*
4. The third feature is the uplift of the corner of the mouth, compared with the middle of the lower lip d1 and d2, as shown in the figure, normalized by the difference of the Y coordinates of point A point B and point D point E, respectively and represented by *lc*.
5. The fourth feature is the angle of line AB and line DE with the central vertical profile, represented by *ag*.
6. The last two features are extracted from the semicircular areas shown, which are defined by using line AB and line DE as diameters. The histograms of the range (Z coordinates) of all the points within these two semicircles are calculated.

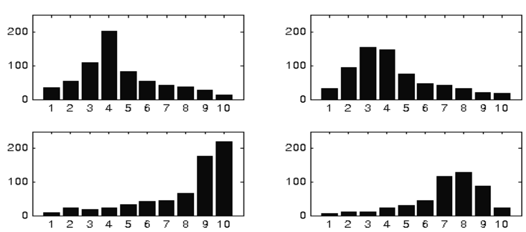

Figure 4 Histogram of range of cheeks (L &R) for neutral (top row), and smiling (bottom row) face

Figure 4 shows the histograms for the smiling and the neutral faces of the subject in Figure 3. The two figures in the first row are the histograms of the range values for the left cheek and right cheek of the neutral face image; the two figures in the second row are the histograms of the range values for the left cheek and right cheek of the smiling face image.

From the above figures, we can see that the range histograms of the neutral and smiling expressions are different. The smiling face tends to have large values at the high end of the histogram because of the bulge of the cheek muscle. On the other hand, a neutral face has large values at the low end of the histogram distribution. Therefore two features can be obtained from the histogram:

One is called the 'histogram ratio', represented by *hr*, the other is called the 'histogram maximum', represented by *hm*.

$$hr = \frac{h6 + h7 + h8 + h9 + h10}{h1 + h2 + h3 + h4 + h5} \quad (1)$$

$$hm = i\,; \quad i = \arg\{\max(h(i))\} \quad (2)$$

After the six features have been extracted, this becomes a general classification problem. Two pattern classification methods are applied to recognize the expression of the incoming faces. The first method used is a linear discriminant (LDA) classifier, which seeks the best set of features to separate the classes. The other method used is a support vector machine (SVM). For our work, Libsvm [7] was used to implement a suitable support vector machine.

5. 3D FACE RECOGNITION

5.1. Neutral face recognition

In our earlier research work, we have found that the central vertical profile and the contour are both discriminant features for every person[8]. Therefore, for neutral face recognition, the same method as in [9] is used: the results of central vertical profile matching and contour matching are combined. The combination of the two classifiers improves the overall performance significantly. The final similarity score for the probe image is the product of ranks for each of the two classifiers (based on the central vertical profile and contour). The image with the smallest score in the gallery will be chosen as the matching face for the probe image.

5.2. Smiling face recognition

For the recognition of smiling faces we have adopted the probabilistic subspace method proposed by B. Moghaddam et al. [10, 11]. It is an unsupervised technique for visual learning, which is based on density estimation in high dimensional spaces using an eigen decomposition. Using the probabilistic subspace method, a multi-class classification problem can be converted into a binary classification problem.

In the experiment for smiling face recognition, because of the limited number of subjects (30), the central vertical profile and the contour are not used directly as vectors in a high dimensional subspace. Instead, they are down sampled to a dimension of 17 to be used. The dimension of difference in feature space is set to be 10, which contains approximately 97% of the total variance. The dimension of difference from feature space is 7.

In this case also, the results of central vertical profile matching and contour matching are combined, improving the overall performance. The final similarity score for the probe image is the product of ranks for each of the two classifiers. The image with the smallest score in the gallery will be chosen as the matching face for the probe image.

6. EXPERIMENTS AND RESULTS

One gallery and three probe databases were used for evaluation. The gallery database has 30 neutral faces, one for each subject, recorded in the first data acquisition session. Three probe sets are formed as follows:

Probe set 1: 30 neutral faces acquired in the second session.
Probe set 2: 30 smiling faces acquired in the second session.
Probe set 3: 60 faces, (probe set 1 and probe set 2).

Experiment 1: Testing the expression recognition module

The leave-one-out cross validation method is used to test the expression recognition classifier. Every time, the faces collected from 29 subjects in both data acquisition sessions are used to train the classifier and the four faces of the remaining subject collected in both sessions are used to test the classifier. Two classifiers are used. One is the linear discriminant classifier; the other is a support vector machine classifier. LDA tries to find the subspace that best discriminates different classes by maximizing the between-class scatter matrix, while minimizing the within-class scatter matrix in the projective subspace. Support vector machine is a relatively new technology for classification. It relies on preprocessing the data to represent patterns in a high dimension, typically much higher than the original feature space. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two categories can always be separated by a hyperplane.

Table 1 Expression recognition results

Method	LDA	SVM
Expression recognition rate	90.8%	92.5%

Experiment 2: Testing the neutral and smiling recognition modules separately

In the first two sub experiments, probe faces are directly fed to the neutral face recognition module. In the third sub experiment, the leave-one-out cross validation is used to verify the performance of the smiling face recognition module.

2.1 Neutral face recognition: probe set 1. (Neutral face recognition module used.)
2.2 Neutral face recognition: probe set 2. (Neutral face recognition module used.)
2.3 Smiling face recognition: probe set 2. (Smiling face recognition module used.)

From Figure 5, it can be seen that when the incoming faces are all neutral, the algorithm which treats all the faces as neutral achieves a very high recognition rate.

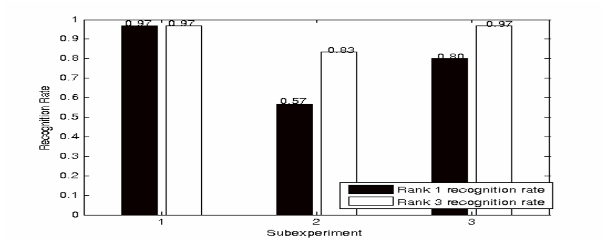

Figure 5 Results of Experiment 2(three sub-experiments)

On the other hand, if the incoming faces are smiling, then the neutral face recognition algorithm does not perform well, only 57% rank one recognition rate is obtained. (Rank one means only the face which scores highest is selected from the gallery. Rank one recognition rate is the ratio between number of faces correctly recognized and the number of probe faces. Rank three means three highest scored faces instead of one face are selected.) In contrast, when the smiling face recognition algorithm is used to deal with smiling faces, the recognition rate can be as high as 80%.

Experiment 3: Testing a practical scenario

These experiments emulate a realistic situation in which a mixture of neutral and smiling faces (probe set 3) must be recognized. Sub experiment 1 investigates the performance obtained if the expression recognition front end is bypassed, and the recognition of all the probe faces is attempted with the neutral face recognition module alone. The last two sub experiments implement the full framework shown in Figure 1. In 3.2 the expression recognition is performed with the linear discrimant classifier, while in 3.3 it is implemented through the support vector machine approach.

3.1 Neutral face recognition module used alone: probe set 3 is used
3.2 Integrated expression and face recognition: probe set 3 is used. (Linear discriminant classifier for expression recognition.)
3.3 Integrated expression and face recognition: probe set 3 is used. (Support vector machine for expression recognition.)

It can been seen in Figure 6 that if the incoming faces include both neutral faces and smiling faces, the recognition rate can be improved about 10 percent by using the integrated framework proposed here.

7. DISCUSSION AND CONCLUSION

The work reported in this paper represents an attempt to acknowledge and account for the presence of expression on 3D face images, towards their improved identification. The method introduced here is computationally efficient. Furthermore, this method also yields as a secondary result the information of the expression found in the faces.

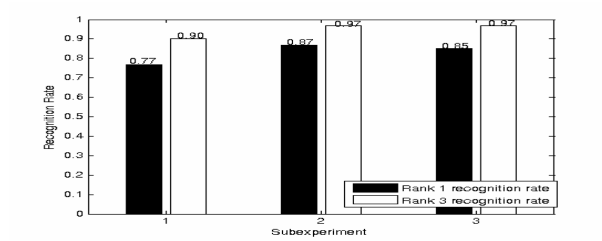

Figure 6 Results of Experiment 3(three sub-experiments)

Based on these findings we believe that the acknowledgement of the impact of expression on 3D face recognition and the development of systems that account for it, such as the framework introduced here, will be keys to future enhancements in the field of 3D Automatic Face Recognition.

8. ACKOWELDGEMENT

This work was sponsored by NSF grants IIS-0308155, CNS-0520811, HRD-0317692 and CNS-0426125. Mr. Chao Li is the recipient of an FIU Dissertation Year Research Fellowship.

9. REFERENCES

[1] R. Chellappa, C.Wilson, and S. Sirohey, “Human and Machine Recognition of Faces: A Survey,” Proceedings of the IEEE, 1995. **83(5)**: pp. 705-740.
[2] W. Zhao, R. Chellappa, and A.Rosenfeld, “Face recognition: a literature survey,” *ACM Computing Survey*, 2003. **35**: pp. 399-458.
[3] K. Bowyer, K. Chang, and P. Flynn, “A Survey of Approaches to 3D and Multi-Modal 3D+2D Face Recognition,” *IEEE Intl. Conf. on Pattern Recognition*, 2004.
[4] N. Etcoff, J. Magee, “Categorical perception of facial expressions,” *Cognition*, 1992. **44**: pp. 227-240.
[5] P.Ekman, W. Friesen, “Constants across cultures in the face and emotion,” *Journal of Personality and Social Psychology*, 1971. **17**(2): pp. 124-129.
[6] www.polhemus.com.
[7] C. Chang, C. Lin, LIBSVM: a library for support vector machines, 2001.
[8] C. Li, A. Barreto, “Profile-Based 3D Face Registration and Recognition,” *Lecture Notes on Computer Science*, 2005. **3506**: pp. 484-494.
[9] C. Li, A.Barreto, J. Zhai and C. Chin. “Exploring Face Recognition Using 3D Profiles and Contours,” *IEEE SoutheastCon 2005*. Fort Lauderdale.
[10]B. Moghaddam, A. Pentlend, “Probabilistic Visual Learning for Object Detection,” *International Conferrence of Computer Vision (ICCV' 95)*, 1995.
[11]B. Moghaddam, A. Pentlend, “Probabilistic Visual Learning for Object Representation,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1997. **19**(7): pp. 696-710.