Integrated 3D Expression Recognition and Face Recognition

Chao Li, Armando Barreto
Electrical & Computer Engineering Department
Florida International University
Miami, Florida, 33174, USA
{cli007, barretoa}@fiu.edu
http://dsplab.eng.fiu.edu

Abstract

Face recognition technology has been a focus both in academia and industry for the last couple of years because of its wide potential applications and its importance to meet the security needs of today's world. Most of the systems developed are based on 2D face recognition technology, which uses pictures for data processing. With the development of 3D imaging technology, 3D face recognition emerges as an alternative to overcome the difficulties inherent with 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

Face recognition, together with fingerprint recognition, speaker recognition, etc., is part of the research area known as 'biometric identification' or 'biometrics', which refers to identifying an individual based on his or her distinguishing characteristics. Face recognition is a particularly compelling biometric approach because it is the one used every day by nearly everyone as the primary means for recognition of other humans. Because of its natural character, face recognition is more acceptable than most other biometric techniques. Face recognition also has the advantage of being noninvasive.

Face recognition has a wide range of potential applications for commercial, security, and forensic purposes. These applications include automated crowd surveillance, access control, credit card authorization, design of human computer interfaces, etc. Especially, the surveillance systems rely on the noninvasiveness of face recognition systems. Face recognition approaches can be divided according to the format of the data acquired into 2D face recognition, 3D face recognition and infrared face recognition modalities. The infrared face recognition is commonly combined with other biometric technologies.

Most of the face recognition attempts that have been made until recently use 2D intensity images by photographic cameras as the data format for processing. This kind of research is called 2D face recognition. Varying levels of success have been achieved in 2D face recognition research. Detailed and comprehensive surveys can be found in [1, 2]. Although 2D face recognition has achieved considerable success, certain problems still exist. Because the 2D face images used not only depend on the face of a subject, but also depend on the imaging factors, such as the environmental illumination and the orientation of the subject. These two sources of variability in the face image often make the 2D face recognition system fail. That is the reason why 3D face recognition is believed to have an advantage over 2D face recognition.

With the development of 3D imaging technology, more and more attention has been directed to 3D face recognition. In [3], Bowyer et al. provide a survey of 3D face recognition technology. Some of the techniques are derived from 2D face recognition, such as PCA used in [4, 5]to extract features from faces. Some of the techniques are unique to 3D face recognition, such as the geometry matching method in [6] and the profile matching proposed in [7, 8]. 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. Therefore, systems that treat the face as a rigid surface are prone to fail when dealing with faces with expressions. In [9], experiments using Iterative Closest Point (ICP) and PCA methods were performed on the recognition of faces with expression. The authors found that expression changes do cause performance to deteriorate in all methods.

Therefore, the involvement of facial expression has become a big challenge in 3D face recognition systems. Up to now, only some methods address the facial expression issue in face recognition. In [10], the authors present a 3D face recognition approach based on a representation of the facial surface , invariant to isometric deformation by facial expression. In [11], both rigid registration and non-rigid deformations caused by expression were calculated. For the purpose of face matching, the non-rigid deformations from different sources are identified, which is formulated as a two-class classification problem: intra-subject deformation vs. inter-subject deformation. The deformation classification results are integrated with the matching distance of rigid registration to make the final decision. In [9], the author tried to extract the area that deforms least with different facial expressions and used this area as the feature for every subject.

In this paper, we tackle the expression challenge in 3D face recognition from a different point of view. Because the deformation of the face surface is always associated with specific expression, an integrated expression recognition and face recognition system is proposed. In section 2, a model of the relationship between expression and face recognition is introduced. Based on this model, the framework of integrated expression recognition and face recognition is proposed. Section 3 explains the acquisition of the experimental data used and preprocessing performed. Section 4 outlines our approach to 3D face expression recognition. Section 5 explains the process used for 3D face recognition. Section 6 describes the experiments and the results obtained. Section 7 is presents our discussion and conclusion.

2. Expression recognition 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. One of the experiments that support the existence of the connection between facial expression recognition and face recognition was reported in [12]. The authors 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 the recognition of faces. 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[13], plus the neutral expression. Therefore, the output of the expression recognition system will be one of the seven expressions. According to different expressions, a matching face recognition system is proposed. If the output is neutral expression, 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 instances of all enrolled subjects, to get the closest match. If the output is other than neutral expression, 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. Since the recognition through modeling is a more complex process than the direct matching for the neutral face, this framework would support the experimental findings, which showed that people are slower in identifying happy and angry faces than they are in identifying faces with neutral expression. Figure 1 shows a simplified version of this framework. This simplified diagram only deals with the happy (smiling) expression, which is the most commonly displayed by people publicly.

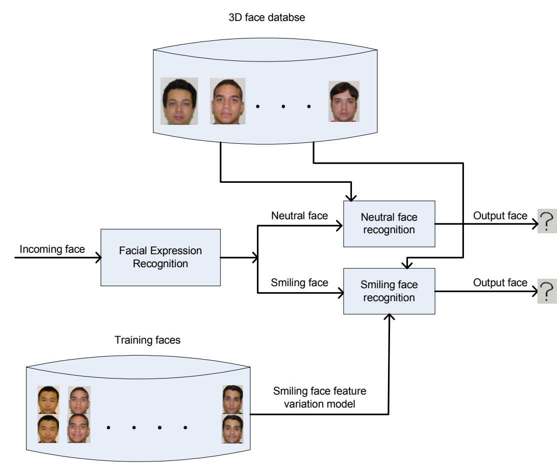

Figure1 Simplified framework of 3D face expression

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 most common expression, 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. One was a neutral expression; the other was a happy (smiling) expression. The 3D scanner used was a Fastscan 3D scanner from Polhemus Inc. [14]. The resulting database contains 60 3D neutral scans and 60 3D smiling scans of 30 subjects. The left image in Figure 2 shows an example of the 3D scans obtained using this scanner.

In 3D face recognition, registration is a key pre-processing step. In our experiment, a method based on the symmetric property of the face is used to register the face image[15]. Because in the scanning process there were some unwanted holes in the face surface, (especially in the area covered by dark hair, such as the eye brows), a cubic spline interpolation method was used to patch the holes. An example of the resulting 3D range images is shown in Fig 2(right image).

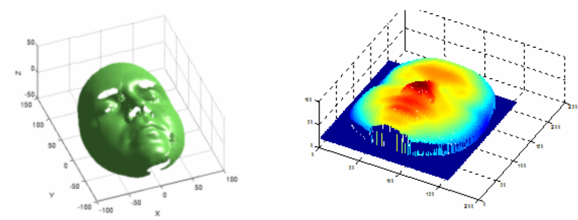

Figure 2 3D surface (left) and mesh plot of the converted range image (right)

4. Expression recognition

The expression of the face is a basic mode of nonverbal communication among people. The facial expressions convey information about emotion, mood and ideas. In [13], Ekman and Friesen proposed six primary emotions. Each possesses a distinctive content together with a unique facial expression. These prototypical emotional displays are also referred to as basic emotions. They seem to be universal across human ethnicities and cultures. These six emotions are happiness, sadness, fear, disgust, surprise and anger. Together with the neutral expression, these expressions also form the seven basic prototypical facial expressions.

Automatic facial expression recognition has gained more and more attention recently. It has various potential applications in improved intelligence in human computer interfaces, image compression and synthetic face animation. "Face expression recognition deals with the classification of facial motion and facial feature deformation into abstract classes that are purely based on visual information." [16].

As in face recognition, most contemporary facial expression systems use two-dimensional images or videos as data format. Logically, the same 2D shortcomings will hamper 2D expression recognition, (i.e., 2D formats are dependent on the pose of the subjects and on the illumination of the environment). In this respect this paper fills the gap by proposing a facial expression system that uses three dimensional images or range images, which are invariant with respect to illumination and subject orientation.

In our experiment, we aim to recognize social smiles, which were posed by each subject. Smiling is the easiest of all expressions to find in photographs and is readily produced by people on demand. It is generated by contraction of the zygomatic major muscle. The zygomatic major originates in the cheek bone (zygomatic arch) and inserts in muscles near the corner of the mouth. 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 bulge of the cheek muscle and the uplift of the corner of the mouth as we can see from the following photos. The line on the face generated by the smiling expression is called the nasal labial fold (smile line).

Figure 3 Illustration of features of smiling face

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

1. An algorithm is developed to obtain the coordinates of five characteristic points A, B, C, D and E in the face range image as shown in Figure 3. A and D are at 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.
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 points BC and EC) normalized by the height of the nose to capture the fact that the smiling expression pulls back the mouth. The 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 points AB and DE, respectively and represented by *lc*.
5. The fourth feature is the angle of AB and 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 AB and DE as diameters. The histograms of the range (Z coordinates) of all the points within these two semicircles are calculated.

The following figure shows the histograms for the smiling face and the neutral face of the above subject.

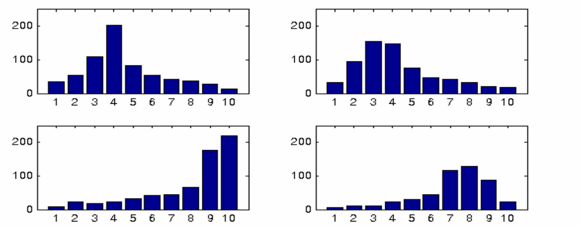

Figure 4 Histogram of range of cheeks for neutral (top row), and smiling (bottom row) face

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

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

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

In summary, six features, i.e. *mw*, *md*, *lc*, *ag, hr* and *hm* are extracted from each face for the purpose of expression recognition.

After the 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 classifier, which seeks the best set of features to separate the classes. The other method used is a support vector machine. 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. [17] For our work, Libsvm [18] 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[15]. Therefore, for neutral face recognition, the same method as in [19] 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. [20, 21]. It is an unsupervised technique for visual learning, which is based on density estimation in high dimensional spaces using an eigen decomposition. The probability density is used to formulate a maximum-likelihood estimation framework for visual search, target detection and automatic object recognition. 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 contain approximately 97% of the total variance. The dimension of difference from feature space is 7.

The results of central vertical profile matching and contour matching are combined. Here also the combination of the two classifiers improves the 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 are formed for the evaluation of our methods in three experiments. 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, constituted by 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. The results shown below are the average of the 30 recognition rates. Two classifiers are used. One is the linear discriminant classifier; the other is a support vector machine classifier.

Table 1 Expression recognition result

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. In each cycle, 29 subjects' faces from both acquisition sessions are used for the training and the remaining subject's smiling face from the session is used as testing face.

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.)

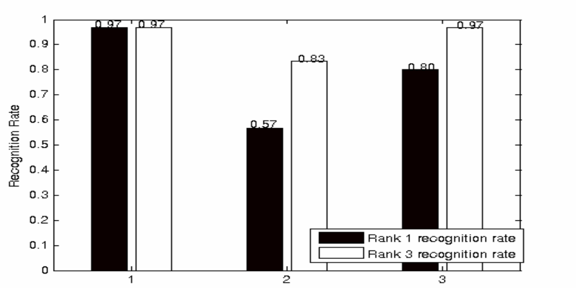

Figure 5 Results of Experiment 2

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. On the other hand, if the incoming faces are smiling faces, then the neutral face recognition algorithm does not perform well, only 57% rank one recognition rate is obtained. 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. (Faces are first sorted according to expression and then routed to the appropriated recognition module.) 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 discriminate classifier for expression recognition.)
3.3 Integrated expression and face recognition: probe set 3 is used. (Support vector machine for expression recognition.)

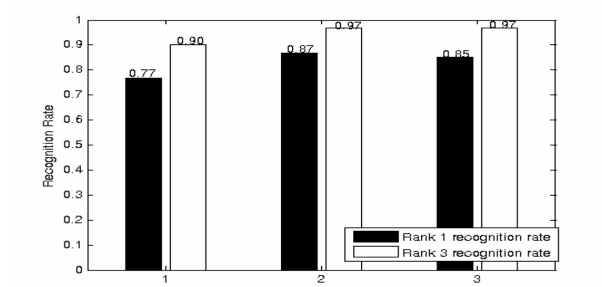

Figure 6 Results of Experiment 3

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

It should be noted that all the experiments involving smiling faces are done using the leave-one-out cross validation method because of the size of the database. Therefore the results displayed are the average, not the best one. For simplicity of implementation, the training samples for the expression recognition system and the smiling face recognition systems are the same faces. In real application, we would select the training samples to make the best classifier for expression recognition and expressional face recognition separately. Considerable performance improvement might be achieved in this way.
Compared with other methods [9-11], this method has the advantage of being computationally efficient. Furthermore, this method also yields the information of the expression found in the faces.

8. Acknowledgement

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, S. Sirohey, Human and machine recognition of faces: a survey. Proceedings of the IEEE, 1995. **83(5)**: p. 705-740.

[2] W. Zhao, R. Chellappa, A. Rosenfeld, Face recognition: a literature survey. ACM Computing Survey, 2003. **35**: p. 399-458.

[3] K. Bowyer, K. Chang, P. Flynn, A Survey of Approaches to 3D and Multi-Modal 3D+2D Face Recognition. IEEE International Conference on Pattern Recognition. 2004: p. 358-361.

[4] K. Chang, K. Bowyer, P. Flynn, Multimodal 2D and 3D biometrics for face recognition. IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003. p. 187-194.

[5] C. Hesher, A. Srivastava, G. Erlebacher, A novel technique for face recognition using range images. Seventh International Symposium on Signal Processing and Its Application, 2003.

[6] G. Gordon, Face recognition based on depth maps and surface curvature. Geometric Methods in Computer Vision, SPIE. July 1991, p.1-12.

[7] J. Cartoux, J. LaPreste, M. Richetin, Face authentication or recognition by profile extraction from range images. Proceedings of the Workshop on Interp. of 3D Scenes. 1989, p.194-199.

[8] T. Nagamine, T. Uemura, I. Masuda, 3D facial image analysis for human identification. Internaitonal Conference on Pattern Recognition (ICPR,1992). 1992, p.324-327.

[9] K. Chang, K. Bowyer, P. Flynn, Effects of facial expression in 3D face recognition. SPIE : Biometric Technology for Human Identification II. 2005.

[10] A. Bronstein, M. Bronsein R. Kimmel, Expression-invariant 3D face recognition. Proc. Audio & Video-based Biometric Person Authentication (AVBPA). 2003,p.62-69.

[11] X. Lu, A. Jain, Deformation analysis for 3D face matching. 7th IEEE Workshop on Applications of Computer Vision (WACV'05). 2005.

[12] N. Etcoff, J. Magee, Categorical perception of facial expressions. Cognition, 1992. **44**: p. 227-240.

[13] P. Ekman, W. Friesen, Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 1971. **17**(2): p. 124-129.

[14] www.polhemus.com.

[15] Chao Li, A. Barreto, Profile-based 3D face registration and recognition. Lecture Notes on Computer Science, 2005. **3506**: p. 484-494.

[16] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey. Pattern Recognition, 2003. **36**: p. 259-275.

[17] R. Duda, P. Hart, D. Stork, Pattern classification. 2nd ed. 2001.

[18] C. Chang, C. Lin, LIBSVM: a library for support vector machines. 2001.

[19] Chao Li, A. Barreto, J. Zhai, C. Chin, Exploring face recognition using 3D profiles and contours. IEEE SoutheastCon 2005. 2005, p.576-579.

[20] B. Moghaddam, A. Pentland, Probabilistic visual learning for object detection. International Conferrence of Computer Vision (ICCV' 95). 1995, p.786-793.

[21] B.Moghaddam, A. Pentlend, Probabilistic visual learning for object representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997. **19**(7): p. 696-710.