Paper 071, ENG 101

Automatic Facial Expression Recognition Using 3D Faces

Chao Li, Antonio Soares
Florida A&M University
chao.li, antonio.soares@famu.edu

Abstract

Automatic facial expression recognition has gained much attention during the last decade because of its potential application in areas such as more engaging human-computer interfaces. Automatic facial expression recognition is a sub-area of face analysis research that is based heavily on methods of computer vision, machine learning, and image processing. Many efforts either to create a novel or to improve existing face expression recognition systems are thus inspired by advances in these related fields. This paper explores the automatic recognition of facial expressions using 3D range images. The paper outlines the development of an algorithm designed to distinguish between neutral and smiling faces, and summarizes its experimental verification with a database containing 30 subjects who posed for both (neutral and smiling) expressions. As a comparison with 2D farcical expression recognition, a PCA algorithm was used to exact features from 2D images and used for expression recognition. Results show that 3D facial expression recognition outperforms 2D ones.

Introduction

In human-to-human dialogue, the articulation and perception of facial expressions form a communication channel that is supplementary to voice and that carries crucial information about the mental, emotional, and even physical states of the conversation partners [1]. As a basic mode of nonverbal communication among people, the facial expression of another person is often the basis on which we form significant opinions on such characteristics as friendliness, trustworthiness, and status. The facial expressions convey information about emotion, mood and ideas.

In [2], Ekman and Friesen proposed six primary emotions. Each emotion 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 seven expressions also form the basic prototypical facial expressions.

Facial expressions are generated by contractions of facial muscles, which result in temporally deformed facial features such as eye lids, eye brows, nose, lips and skin textures, often revealed by wrinkles and bulges. Typical changes of muscular activities for spontaneous expressions are brief, usually between 250ms and 5s. Three stages have been defined for each expression, which are onset (attack), apex (sustain) and offset (relaxation). In contrast to these spontaneous expressions, posed or deliberate expressions can be found very commonly in social interactions. These expressions typically last longer than spontaneous expressions.

Automatic facial expression recognition has gained more and more attention recently. Face expression recognition deals with the classification of facial motion and facial feature deformation into abstract classes that are purely based on visual information.[3] It has various potential applications in improved intelligence for human computer interface, image compression and synthetic face animation. In [4], automatic face recognition is used to build an intelligent tutoring system. In [5] facial expression recognition is used as one way to detect drowsiness of the driver to prevent car accident.

Currently, all existing face expression analysis and recognition systems rely primarily on static images or dynamic videos. A number of techniques were successfully developed using 2D static images or video sequences, including machine vision techniques. [6,7,8]. Although some systems have been successful, the performance degradation remains when handling expressions with large head rotation, subtle skin movement, and/or lighting change with varying postures [9]. Recently, with the development of 3D imaging technology, fast and cheap 3D scanners became available in the market. 3D scans do not have the inherent problems cited above for 2D images. Therefore the extraction of features from the faces is expected to be more robust, which will make the final expression recognition more reliable. In our research 3D range images are used to assess the practicability of 3D facial expression recognition.

In this paper, one specific facial expression, social smile, is used to test our 3D expression recognition system. In our experiment, we sought to recognize social smiles, which were posed by each subject, in their apex period. Smiling is the easiest of all expressions to find in photographs and is readily produced by people on demand. 3D ranges images were used for smiling recognition. In order to compare 3D facial expression recognition with 2D facial expression recognition, a 2D facial recognition algorithm is also employed for the database.

Data Acquisition and Processing

A database including images from 30 subjects was built. In this database, we included faces with smiling, as well as neutral faces from the same subjects. Each subject participated in two data acquisition sessions, which took place in two different days. In each session, two 3D scans were acquired. One was neutral expression; the other was a happy (smiling) expression. At the same time, 2D images were also obtained from the same subjects. The 3D scanner used was a Fastscan 3D scanner from Polhemus Inc [10]. The accuracy of this scanner is specified as 1mm. The resulting database contains 60 3D neutral scans and 60 3D smiling scans of 30 subjects. There are also corresponding 2D images for each 3D scan. Figure 1 shows an example of the 3D scans obtained using this scanner.

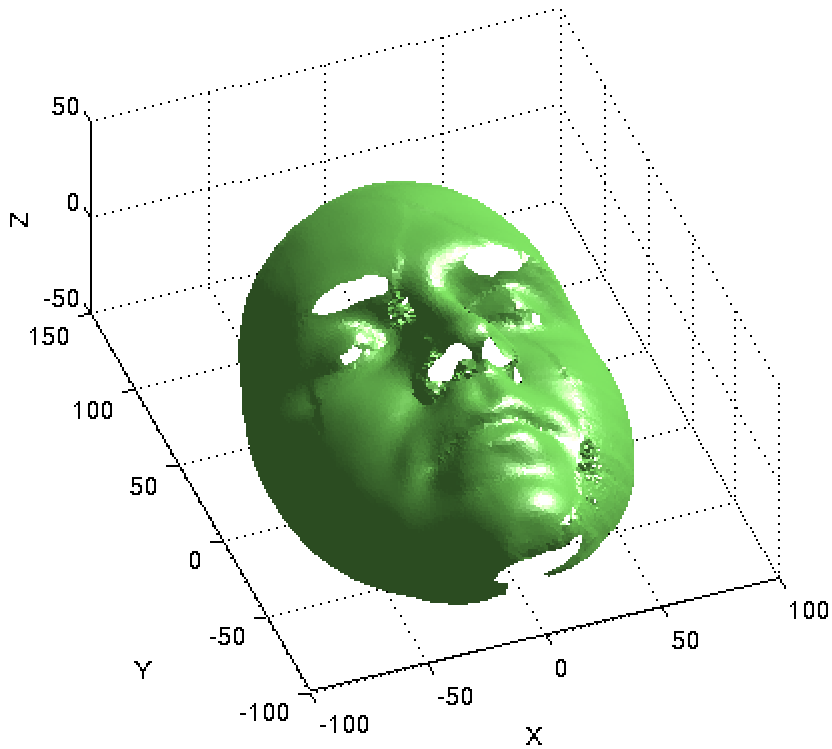

Figure 1. 3D face surface acquired by the 3D scanner

In 3D facial expression 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. In converting the 3D scan from triangulated mesh format to a range image with a sampling interval of 2.5 mm, trilinear interpolation was used [11]. Unavoidably, the scanning process will result in face surfaces containing unwanted holes, especially in the area covered by dark hair, such as the eye brows. To circumvent this problem, the cubic spline interpolation method was used to patch the holes[11]. An example of the resulting 3D range image is shown in Fig 2.

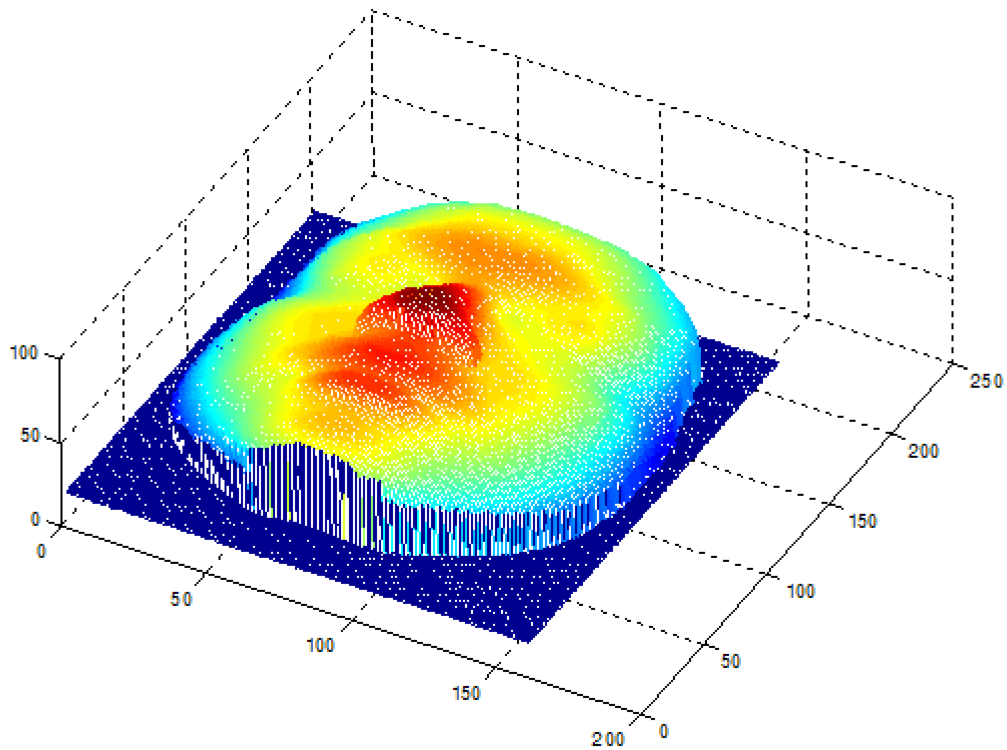

Figure 2. Mesh plot of the converted range image

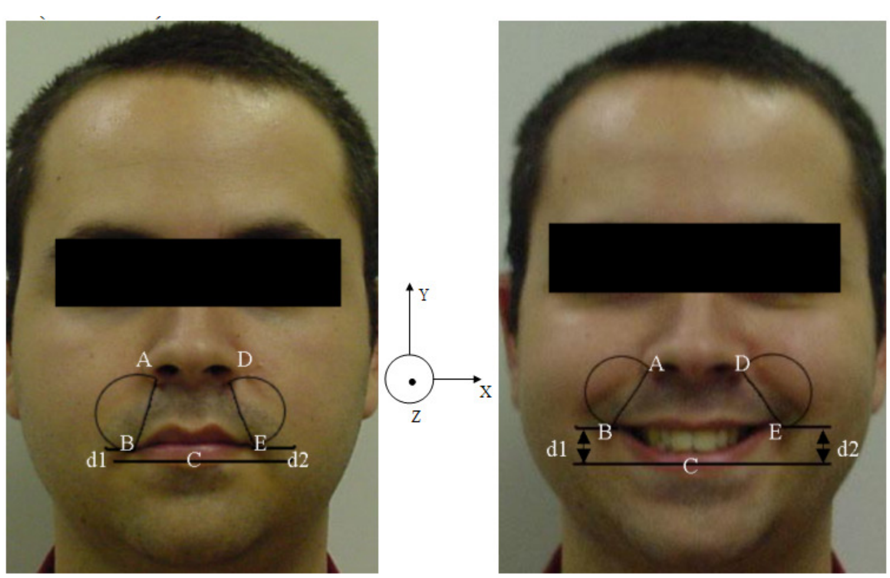

Figure 3. Illustration of the features of a smiling face

Feature Extract action and Classification

The smile is generated by the 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 a smile are the bulge of the cheek muscle and the uplift of the corner of the mouth, as can be seen in Fig 3. The line on the face generated by a smiling expression is called the nasal labial fold (smile line).

The following steps are followed to extract the features for the smiling expression from a 3D range facial image:

- 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 Figure3. 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.
- 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*.
- 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. This second feature is represented by *md.*
- 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 Figure1, normalized by the difference of the Y coordinates of points AB and DE, respectively and represented by *lc*.
- The fourth feature is the angle of AB and DE with the central vertical profile, represented by *ag*.
- The last two features are extracted from the semicircular areas, 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.

Figure 4 shows the histograms for the smiling face and the neutral face of the subject shown in Figure 3.

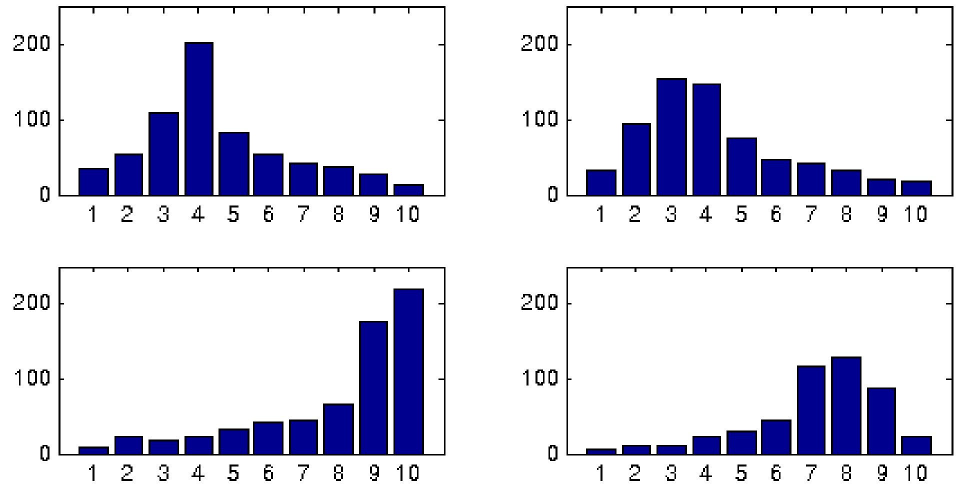

Figure 4. Histogram of range of cheeks for neutral and smiling face

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 obtained from the histograms:

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

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

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

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.

1. Linear discriminant classifier: (Linear Discriminant Analysis-LDA)

LDA tries to find the subspace that best discriminates different classes by maximizing the between-class scatter matrix S_b, while minimizing the within-class scatter matrix S_w in the projective subspace. S_w and S_b, are defined as follows,

$$S_w = \sum_{i=1}^{L} \sum_{\vec{x}_k \in X_i} (\vec{x}_k - \vec{m}_i)(\vec{x}_k - \vec{m}_i)^T \quad (3)$$

$$S_b = \sum_{i=1}^{L} n_i (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T \quad (4)$$

Where $\vec{m}_i$ is the mean vector for the individual class and n_i is the number of samples in class X_i, $\vec{m}$ is the mean vector of all the samples. L is the number of classes.
The LDA subspace is spanned by a set of vectors W, satisfying

$$W = \arg\max \left| \frac{W^T S_b W}{W^T S_w W} \right| \quad (5)$$

2. Support Vector Machine (SVM):

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 [12]. In our research, the Libsvm program package [13] was used to implement the support vector machine.

In order to compare the 3D facial expression algorithm with 2D facial expression algorithm, the corresponding 2D images were used to for expression recognition. First, 2D images were cropped to just keep the face part, eliminating the hair and other artifacts in the 2D image. Then instead of exacting features from 2D images intuitively as in 3D face expression recognition does, Principal Component Analysis (PCA) is used to extract the “feature” from 2D images[14].

PCA

PCA seeks a projection that best represents the data in a least-square sense. In PCA, a set of vectors are computed from the eigenvectors of the sample covariance matrix C,

$$C = \sum_{i=1}^{M} (\vec{x}_i - \vec{m})(\vec{x}_i - \vec{m})^T \quad (6)$$

where $\vec{m}$ is the mean vector of the sample set. The eigen space Y is spanned by k eigenvectors $u_1, u_2 \dots \dots u_k$, corresponding to the k largest eigen values of the covariance matrix C.

$$\vec{y}_i = (\vec{x}_i - \vec{m})^T [\vec{u}_1 \vec{u}_2 \dots \vec{u}_k] \tag{7}$$

The dimensionality of vector $\vec{y}_i$ is K *(K<<M).*

These K eigen valuses are served as the "features" in 2D images. Then the same LDA and SVM methods are used for facial expression recognition.

Experiments and Results

Because the size of the database is relatively small, the leave-one-out cross validation method is used to test the facial expression recognition algorithm. The images of 29 subjects are used to train the classifier, which is used to recognize the expression of the one remaining subject. The results of recognition hits shown below are correct expression recognition (either neutral or smiling), divided by the total number of trials.

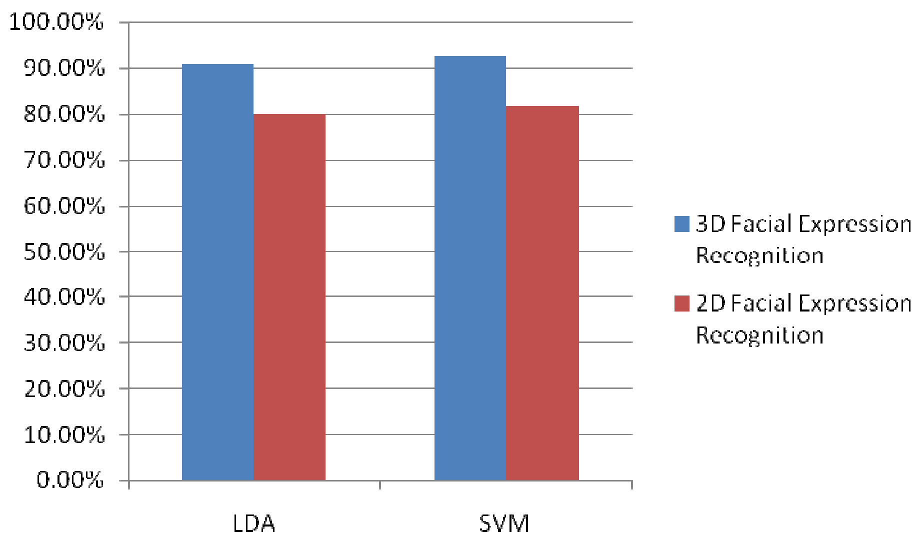

Figure 5. Facial Expression Recogniton Result

Discussion and Conclusion

From Figure 5, it can be seen both classifiers (LDA & SVM) achieve very good facial expression recognition rates for 3D images; both are more than 90%. Otherwise for 2D image, the recognition for both classifiers is around 80%. 3D images have achieved significant better recognition rate than 2D images. This result is in line with our assumption that because of the advantages of 3D images, 3D facial expression recognition system should perform better than its 2D counterpart.

It should also be noted that this experiment, as implemented, pursues the recognition of "absolute facial expressions". This means that the recognition is being attempted without prior knowledge about the neutral facial expression of a subject. It is always more difficult to recognize absolute facial expressions, without referring to the neutral face of a given subject. In many real scenarios, we could incorporate the knowledge of the neutral expression of a subject and modify the algorithm to achieve better performance.

References

[1] Whitehill, J., "Automatic Real-Time Facial Expresssion Recognition for Signed Language Translation", Master thesis, University of West Cape, South Afirca, 2006.

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

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

[4] Whitehill,J, Bartlett,M and Movellan,J, "Automatic Facial Expression Recognition for Intelligent Tutoring Systems", Proceedings of the CVPR Workshop on Human Communicative Behavior Analysis, 2008.

[5] Vural,E, Cetin,M, Ercil, A, Littlewort, G,Bartlett, M and Movellan,J ,"Automated drowsiness detection for improved driving safety". Proc. Int'l Conf. Automotive Technologies. 2008.

[6] Donato,G, Bartlett, M, Hager,J, Ekman, P and Sejnowski, T, "Classifying facial actions", IEEE Trans. PAMI, 21(10):974-989, 1999.

[7] Lyons, M et al. "Automatic classification of single facial images", IEEE Trans. PAMI, 21(12):1357-1362, 1999.

[8] Essa I, and Pentland, A. "Coding, analysis, interpretation, and recognition of facial expressions", IEEE Trans. PAMI, 19(7), 1997.

[9] Yin, L. et al. "A 3D Facial Expression Database For Facial Behavior Research" Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, FGR06, Southampton, UK, 10-12 April 2006

[10] http://www.polhemus.com

[11] Li, C., Barreto, A., "Profile-Based 3D Face Registration and Recognition", Lecture Notes on Computer Science **3506**: 484-494. 2005

[12] Duda, R., Hart, P., Stork D., Pattern Classification, 2001.

[13] Chang, C and Lin, C. *LIBSVM: a library for support vector machines*. http://www.csie.ntu.edu.tw/~cjlin/libsvm/, 2001.
[14] Turk, M and Pentland, A. "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol 3, pp71-86, 1991.

Biography

CHAO LI is currently an Assistant Professor in Electronic Engineering Technology Program at Florida A&M University. He got his PhD in Electrical Engineering from Florida International University. He worked for telecommunication and electronics industry for five years before entering academia.

ANOTONIO SOARES LI is currently an Assistant Professor in Electronic Engineering Technology Program at Florida A&M University. He received a Bachelor of Science degree in Electrical Engineering from Florida Agricultural and Mechanical University in Tallahassee (FAMU), Florida in December 1998. Then he got a Master of Science degree in Electrical Engineering from FAMU in December of 2000 with focus on semiconductor devices, semiconductor physics, Optoelectronics and Integrated Circuit Design. He then worked for Medtronic as a full-time Integrated Circuit Designer for several years before coming back to FAMU to pursue his PhD degree, which he completed in 2008.