3D Facial Expression Recognition for the Enhancement of Human-Computer Interaction

Chao Li, Armando Barreto

Electrical & Computer Engineering Department
Florida International University
Miami, FL 33174
{cli007, barretoa}@fiu.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. 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. The results of the experiments are very encouraging.

Introduction

In today's world, computers have become an essential partner in people's daily life. But the interaction between computers and human beings is still unnatural. In an effort to achieve a more engaging human computer interaction, our research explores the use of facial expression recognition to enhance the human-computer exchange. Facial expression is a basic mode of nonverbal communication among people. The facial expressions convey information about emotion, mood and ideas. In [Ekman and Friesen, 1971], Ekman and Friesen proposed six primary emotions, which are happiness, sadness, fear, disgust, surprise and anger. Each possesses a distinctive content together with a unique facial expression. Together with the neutral expression, these expressions also form the basic prototypical facial expressions. Most contemporary facial expression recognition systems use two-dimensional images or videos as data format, which are dependent on the pose of the subjects and are sensitive to the illumination of the environment. 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.

Data Acquisition and Preprocessing

To test the performance of the facial expression recognition algorithm, a database including 3D scans from 30 subjects was built. Each subject participated in two data acquisition sessions. In each session, two 3D scans were acquired. One was a neutral expression; the other was a happy (smiling) expression. Some preprocessing steps, including registration and interpolation were applied to the scanned images. The left image in Figure 1 is an example of the resulting 3D range image used in the experiment.
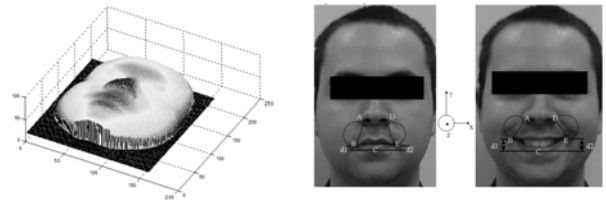

Fig1 (Left) Mesh plot of the converted range image; (Right) illustrations of the features of a smiling face

Feature Extract action and Classification

In our experiment, we sought to recognize smiles. Smiling is the easiest of all expressions to find in photographs and in everyday life. The most distinctive features associated with a smile are the bulge of the cheek muscles and the uplift of the corners of the mouth, as can be seen in the right panel of Figure1. 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 Figure1. 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 2 shows the histograms for the smiling face and the neutral face of the subject shown in Figure 1.
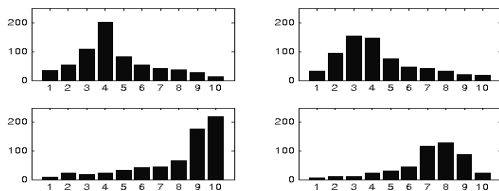

Fig 2 Histogram of range values of face cheeks

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. A smiling face tends to have large values at the high end of the histogram because of the bulge of the cheek muscle, while a neutral face is the opposite. Therefore two features can be obtained from the histograms: One is called the 'histogram ratio', represented by *hr*, and the other is called the 'histogram maximum', represented by *hm*:

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

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

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 faces in the database. One is Linear Discriminant Classifier (LDA); the other is Support Vector Machine Classifier. LDA is used to try 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 machines rely 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. In our research, the Libsvm program package [Chang and Lin, 2001] was used to implement the support vector machine.

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 results shown below are the average of the 30 recognition outcomes.

Table 1 Facial expression recognition rate

Method	LDA	SVM
Expression recognition rate	90.8%	92.5%

Discussion and Conclusion

From Table 1, it can be seen both classifiers achieve a very good facial expression recognition rate. 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.

This experiment has verified the practicability of using 3D range images in automatic facial recognition for the classification of two types of expression: neutral and smiling. Future research will explore the extension of the method used here to the identification of other types of expressions.

Acknowledgement

References

Chang, C and Lin, C. 2001. *LIBSVM: a library for support vector machines.*
http://www.csie.ntu.edu.tw/~cjlin/libsvm/

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