I-069

Smile and Laugh Recognition from Natural Conversation Video

Wang Xinyue † Motoyuki Suzuki † Akinori Ito † and Shozo Makino †

1. Introduction

Tools for human facial expression analysis are indispensable for the realization of human-computer interfaces. Recently, much interest has been shown in developing a human facial expression recognition system[1][2]. It is difficult to recognize a person's facial expressions during conversation, because the lips and face muscles are constantly moving.

Since laughter is observed very often during human communication, we propose a simple and effective method for recognizing the laugh state in a natural conversation. The movement of the face muscles and the feature points on the lips and eyes are observed to realize the laugh state recognition over a natural conversation facial video.

2. Natural Conversation Database

In many applications of human-computer interaction, it is important to detect the emotional state of the person in a natural situation. However, getting a real smile can be challenging. Hence, future human-computer communication systems will be required to recognize a person's natural facial expression during conversation. Previous works dealt with still images or short video sequences of artificial facial expressions. In our research, we focus on long video sequences of natural conversations. To this end, we have constructed a database of natural conversations, which is composed of natural conversations by 7 persons with 3 kind of languages. Figure 1 shows how visual and audio data are recorded for our natural conversation database.

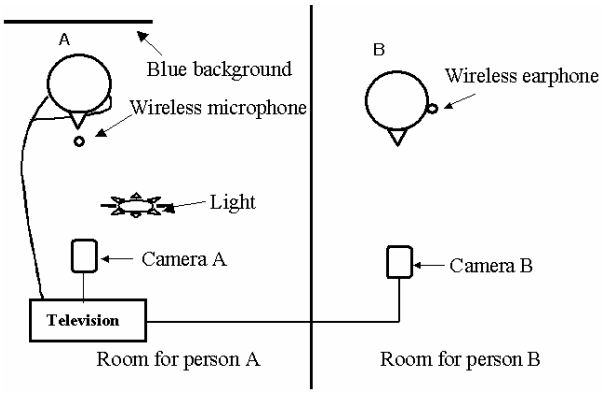

Figure 1: The settings to build the database

We had two persons in two different rooms because we wanted to record person A's voice data for further research. Person A and person B talked freely with any topic that they were interested in. We recorded the visual and audio data of person A for 6 minutes to 8 minutes.

3. Facial Expression Recognition System

3.1The Recognition method

Our recognition method is based on facial feature point extraction and classification of feature vectors. We choose this method because of its simplicity and ease of implementation. In this paper, a 6-dimensional feature vector is constructed for each video frame, and a linear discriminant function is used to classify the 6D feature vectors into two different categories, which we define as laugh state or otherwise. Figure 2 illustrates the overall structure of our proposed system.

In order to construct the 6D feature vector, we extract the following feature components: lip angle θ_1, θ_2, lip length L_1, L_2, and cheek shadows C_1, C_2. The cheek shadows C_1, C_2 are defined as the mean of the gray pixel values of the rectangular windows W_1 and W_2, which correspond to the position of the left and right cheek, respectively. The coordinates of W_1 and W_2 are derived from the position of nose N and upper lip U. We combine these six feature components into a 6-dimensional feature vector **V**, where $\mathbf{V}=(\theta_1,\theta_2,L_1,L_2,C_1,C_2)$. An illustration of the feature vector components is shown in Figure 3.

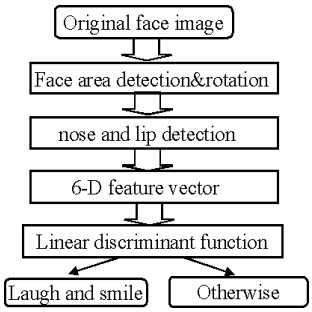

Figure2.System diagram.

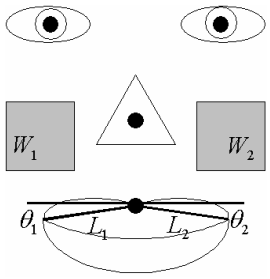

Figure 3: Extraction of feature vector components.

3.2 Feature Extraction

In order to get the 6-D feature vector, robust, accurate extraction of facial features is essential.. Here, we describe how to extract meaningful feature information from a face image for recognition

3.2.1 Facial Area Detection and Rotation

Color is a distinctive feature of the human face. hence it is useful for locating a human face on still and video images. We segment the face region from the video image by detecting the skin color. Given an RGB color image I(x,y)=(*R*,*G*,*B*), a *skin mask* is constructed using pixels that satisfy all of the following

† Graduate School of Engineering, Tohoku Univ.
† {wxinyue, moto ,aito,makino}@makino.ecei.tohoku.ac.jp

criteria [3].

$$0.333 < r < 0.664,\ r > g,$$
$$0.246 < g < 0.398,\ g > 0.5 - 0.5\ r,$$

where r = R/(R+G+B), and g=G/(R+G+B). Median filtering is employed to remove segmentation noise. A example of a skin mask is shown in Figure 4(b).

Since the head may tilt slightly to the left or right side during natural conversation, it is necessary to rotate the image to the proper orientation for further feature extraction. The positions of the eyes are used to estimate the angle of tilt.

To locate the eyes, we first convert the RGB color image to grayscale, then we search for the darkest image blocks in the upper left and upper right region of the skin mask. The position of the left eye and right eye are given by the centers of the two blocks. To account for the shape of the eye, a horizontal thin block is used. This is to avoid false detection of the eyebrow or nostril as the eye.

3.2.2 Detection of Nose and Lip

In order to detect the position of the nose and lip, a horizontal Sobel filter is applied to extract the horizontal edges from the grayscale image. The resultant edge image is thresholded into a black-and-white image. Connected component analysis is applied to remove small regions. Finally, the position of the nose *N* and upper lip *U* are detected by scanning down from the midpoint of the eyes.

An example of feature extraction is given in Figure 4(c).

(a) (b) (c)

Figure 4: Feature Extraction: (a) original image, (b) skin mask, and (c) extracted features.

4. Experiment

In our experiment, we investigate the recognition capability of our proposed system by analyzing a natural conversation video sequence of the test subject shown in Figure 4(a). Feature extraction described in Section 3 was done for each frame to construct the 6-dimensional feature vectors. We used 30 seconds of video for supervised learning, where the correct classification (laugh state or otherwise) is provided with human assistance. We then use another 90 seconds of video for classification after learning. The results show that the smile and laugh can be recognized effectively. The frame rate is 30 fps. Some examples of face images from this video sequence are shown in Figure 5.

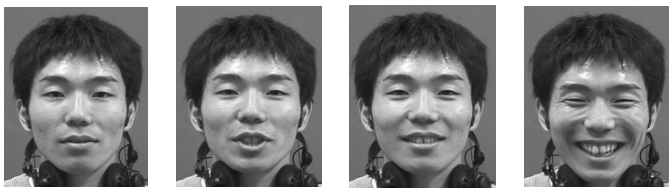

Figure 5: Examples of images from natural conversation video. (From left to right: peace, speaking, smile, laughter)

The classification result compared with human classification is shown in Table 1. We observe that the recognition rate reflects the trend of actual changes in the person's facial expression. However, the system occasionally fails to detect the laugh state when the subject is both laughing and talking at the same time.

Table 1: Laugh and smile state recognition result of proposed system. (The number of frames is shown in parentheses.)

Human/System	Laugh	Otherwise
Laugh (1225)	83.4% (1022)	16.6% (203)
Otherwise (1475)	1.6% (24)	98.4% (1451)

5. Conclusion

In this paper, we proposed a facial expression recognition system for recognizing the smile and laugh state from natural conversation video. In contrast to previous works which dealt with still images or short video sequences of artificial expressions, our work focuses on analyzing long video sequences stored in a natural conversation database. Our method uses feature extraction and classification of feature vectors by a linear discriminant function.

Further extraction of facial features for more robust analysis, and comparison with other recognition methods are left for future study. We also plan to include audio data for multimodal analysis of facial expression to improve the recognition rate.

References

[1]Y. Tian, T. Kanade and J.F. Cohn. "Recognizing Actin Units for Facial Expression Analysis". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.23, No.2 February, 2001.

[2]Yaser Yacoob and Larry S. Dabis. "Recognizing Human Facial Expressions From Long Image Sequences Using Optical Flow". *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.18, No.6 June, 1996.*

[3] Yuiti Araki,, Nobutaka Shimada, and Yoshiaki. "Detection of Faces of Various Direction and Estimation of Face" Direction in Complex Backgrounds. *Technical Report of IEICE PRMU* 2001-217 pp.87-94 (2002-01)