A System for Facial Expression Analysis of a Person While Using Video Phone

Taro Asada, Yasunari Yoshitomi, Ryota Kato, and Masayoshi Tabuse
Graduate School of Life and Environmental Sciences, Kyoto Prefectural University,
1-5 Nakaragi-cho, Shimogamo, Sakyo-ku, Kyoto 606-8522, Japan
E-mail: {t_asada, r_kato}@mei.kpu.ac.jp, {yoshitomi, tabuse}@kpu.ac.jp

Jin Narumoto
Graduate School of Medical Science, Kyoto Prefectural University of Medicine,
Kajii-cho, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto 602-8566, Japan
E-mail: jnaru@koto.kpu-m.ac.jp

Abstract

We developed a method for analyzing the facial expressions of a person having a conversation on a videophone. This was implemented on a wireless local area network, consisting of two personal computers and a router; the subject and the operator are placed in front of each computer. The module contains a function for determining the reference frame and for locally searching for the mouth area in the image to determine the most appropriate position for each frame.

Keywords: Facial expression analysis, Movement analysis, Mouth area, OpenCV, and Skype.

1. Introduction

In Japan, the average age of the population has been increasing, and this trend is expected to continue. Along with this, the number of older people with dementia and/or depression living in rural areas is increasing very rapidly. Due to a mismatch between the number of patients and the number of healthcare professionals, it is difficult to provide adequate psychological assessment and support for all patients.

To improve the quality of life (QOL) of elderly people living in a care facility or at home, we have developed a method for analyzing the facial expressions of a person who is having a conversation with another person on a videophone.[1, 2] The video is analyzed using image-processing software (OpenCV)[3] and a previously proposed feature parameter (*facial expression intensity*) that is based on the mouth area.[1, 2]

In the present study, exploiting our reported researches[1, 2], we developed a system for facial expression analysis of a person while using video phone.

2. Proposed System and Method

2.1. *System overview and outline of the method*

The platform includes Skype[4] for the videophone, and conversations are recorded for the analysis of facial expressions. In the recorded data, the size of the faces are standardized, and the data are analyzed by using OpenCV and the proposed feature parameters for facial expressions, as outlined below. The Y component obtained from each frame in the dynamic image is used for analyzing the facial expressions. The proposed method consists of (1) the size of the lower part of the face is standardized; (2) the mouth area is extracted; (3)

the facial expression intensity is measured; (4) the reference frame is selected; (5) the best position for mouth area in the frame is determined; (6) utterances are evaluated; and (7) the feature parameter for facial expression strength is calculated. In the following subsections, (2), (3), (4), (5) and (7) are explained in detail. On the details for (1) and (6), see Refs 1-2.

The structure of the proposed system is shown in Fig.1. We developed a wireless local area network, consisting of two personal computers (PCs) and one router, in which the subject is placed in front of a PC with a module for analyzing facial expressions, and the subject converses via videophone with the module operator, who is placed in front of the other PC.

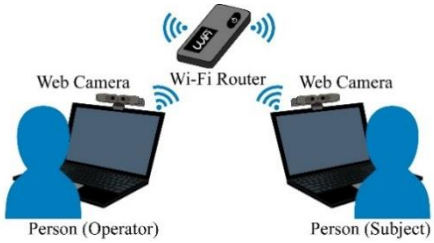


Fig. 1. Structure of the proposed system.

2.2. *Extraction of the mouth area*

Next, by using OpenCV, the mouth area is extracted as a rectangular shape.[1] The mouth area is selected because it is where the difference between neutral and happy facial expressions is most distinct. An example of a face image and the extracted image of the mouth area is shown in Fig. 2.

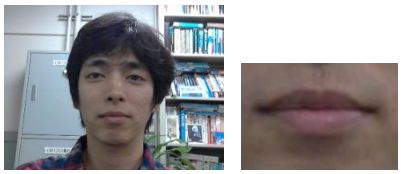


Fig. 2. Whole-face image (left), extracted image of the mouth area (right).

2.3. *Measurement of facial expression intensity*

For the Y component of the selected frame, the feature vector for the facial expression is extracted for the mouth area; this is performed by using a 2D-DCT for each section of 8×8 pixels.

As the feature parameters for expressing facial expression, we selected 15 low-frequency components from the 2D-DCT coefficients; this did not include the direct current component. Next, we obtained the mean of the absolute value of each of these components in the area of the mouth. In total, we obtained 15 values as elements of the feature vector. The facial expression intensity is defined as the norm of the difference vector, which is a vector of the difference between the feature vector for the neutral facial expression and that for the observed expression, and it can be used to analyze changes in facial expression.

The candidate of facial expression intensity defined as the norm of the difference vector between two feature vectors is used for selection of the reference frame as described in the section 2.4.

2.4. *Selection of the reference frame*

We propose a method for automatically selecting the reference frame to be used for measuring the facial expression intensity; this can be used for a moving image in which the initial facial expression is neutral and there is no utterance. The existence of an utterance is determined by a method that we proposed in a previous paper.[1, 2] The frame prior to any utterance is used as the target in following procedure.

We calculated the sum of the norms of the difference vectors for the feature vector calculated for each of the candidates for the reference frame, and we designated the one for which this sum was minimized as the reference frame. The facial expression intensity is then obtained as the norm of the vector of the differences between the feature vector and the feature vector of the reference frame.

2.5. *Correction of position of mouth area*

There is no guarantee that the center of the extracted mouth area will remain in a suitable and constant position by eyes, and in fact, we observed that it did vary. This change in position could affect the evaluated facial expression intensity. Therefore, we propose a way to decrease the influence on the facial expression intensity due to changes in the location of the mouth area.

After using OpenCV to extract the mouth area, the best position of the mouth area under the condition on the shift of the area is found by deciding each shift-value for the area in the range of -3 to 3 pixels in each of the horizontal and vertical directions, in order to give the smallest value of facial expression intensity for the frame.

2.6. *Feature parameter for facial expression strength*

In diagnosing a patient having dementia and/or depression, it might be useful for healthcare professionals to evaluate the strength of facial expressions by using a simple measure.[1] Moreover, it might be more advantageous for a diagnosis of dementia and/or depression to separately evaluate the strength of the facial expression as a speaker and a listener.[1] Therefore, we measure the feature parameter for facial expression strength as the average of facial expression intensity as a speaker and a listener[1], if necessary.

3. Experiments

3.1. *Experimental environment*

The experiment was performed in the following computational environment: the PC set in front of subject A was a Dell Vostro 3350; CPU: Intel Core i7-2620M 2.7 GHz; 4.0 GB memory; OS: Microsoft Windows 7 Professional. The development language was Microsoft Visual C++ 2008 Express Edition.

Three males (subject A in his 30s, subject B in his 20s, and subject C in his 40s; C is a psychiatrist) participated in addition to an operator. Each experiment consisted of a two-way videophone conversation that lasted from 27 to 36 seconds. Experiment 1 was a conversation between subject A in Tokyo and subject B in Kyoto, and it was conducted using Skype. Experiment 2 took place face-to-face as an interview to subject A by subject C. We saved the visual and audio information as AVI files, and these were then used for measuring the feature parameters of the facial expressions. The size of the image frame was 640×480 pixels, and the size of the standardized lower part of the face image was set to 240×96 pixels.

3.2. *Results and discussion*

Fig. 3 shows the mouth area of subject A at the beginning of each experiment.

Fig. 3. The mouth area of subject A at the beginning of Experiment 1 (left), and at the beginning of Experiment 2 (right).

The facial expression intensity for subject A was recorded during both experiments. In Experiment 1, the timing of utterances was 11.4 seconds. During his utterances in Experiment 1, there were four local peaks in the facial expression intensity of subject A. These occurred at approximately 3, 6, 15, and 21 seconds after the starting point (see Fig. 4). During the times when subject A was not making utterances during Experiment 1, there were five local peaks in his facial expression intensity; these were at approximately 3, 7, 11, 13, and 23 seconds after the starting point (see Fig. 5); characteristic images of the face and mouth area are shown in Fig.5. In Experiment 2, the timing of utterances was 12.4 seconds; there were four local peaks in the facial expression intensity for subject A; these were at approximately 14, 16, 23, and 26 seconds after the starting point (see Fig. 6). In both Experiments 1 and 2, the images of the face and mouth areas at the characteristic timing points show that the proposed method can quantitatively determine the facial expression (see Figs. 5 and 6).

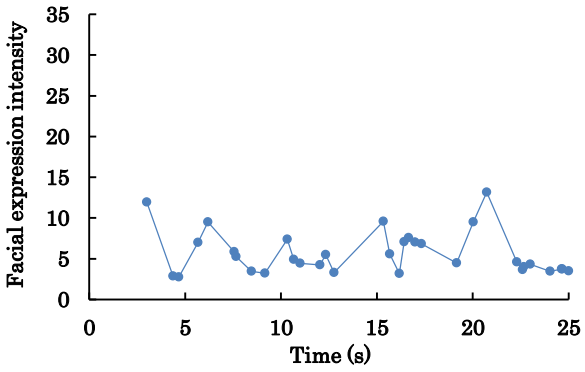
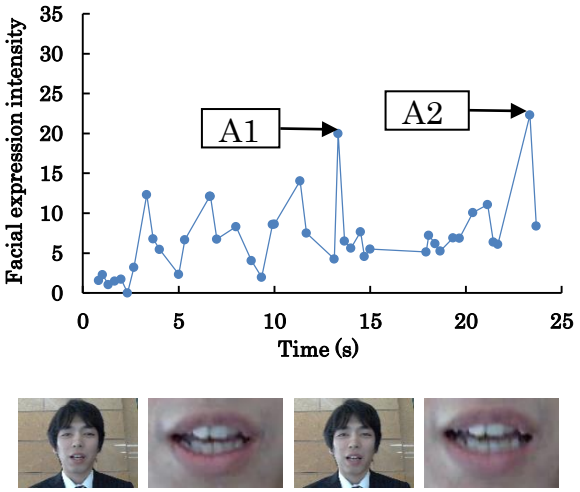
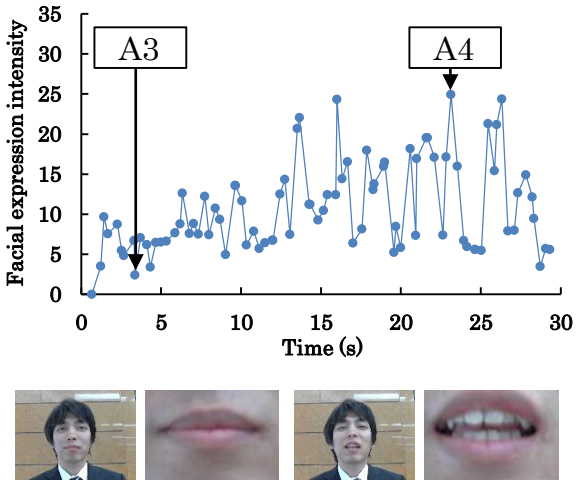


Fig. 4. Changes in the facial expression intensity for subject A at times of utterance during Experiment 1.



A1 A2

Fig. 5. Changes in the facial expression intensity for subject A at times of no utterance during Experiment 1 (upper graph). Whole-face images and mouth images are shown for two moments, A1 and A2, which are indicated on the graph (lower images).



A3 A4

Fig. 6. Changes in the facial expression intensity for subject A during Experiment 2 (upper graph). Whole-face images and mouth images are shown for two moments, A3 and A4, which are indicated on the graph (lower images).

The feature parameters for the facial expression intensity are shown in Table 1. Note that the intensity tended to be greater during face-to-face conversation than during a videophone conversation.

The processing time for analyzing the facial expressions using the proposed system was 126 s for Experiment 1 and 135 s for Experiment 2.

Table 1. Feature parameter for facial expressions of subject A.

Experiment	Utterance	Feature parameter
1	with	5.65
	without	6.96
2	with & without	10.67

4. Conclusion

We developed a wireless local area network system, consisting of two PCs and one router, which includes a module for analyzing the facial expressions of a subject engaged in a two-way videophone conversation.

The results show the usefulness of the proposed method. As an area of future work, we intend to develop a related method for estimating the mental state and/or recognition ability of a patient.

Acknowledgements

This research was supported by COI STREAM of the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1. T. Asada, Y. Yoshitomi, R. Kato, M. Tabuse, and J. Narumoto, Quantitative evaluation of facial expressions and movements of persons while using video phone, *J. Robotics, Networking and Artif. Life* **2**(2) (2015) 111-114.
2. T. Asada, Y. Yoshitomi, R. Kato, M. Tabuse and J. Narumoto, Analysis of facial expressions robust against small imperfections in mouth-part area extraction from face-images of persons while using video phone (in Japanese), in *Proc. of Human Interface Symposium 2015* (Japan, Hakodate, 2015), pp.187-190.
3. Open CV. http://opencv.org/ Accessed 1 December 2015.
4. Skype Web page, http://www.skype.com/ Accessed 5 November 2015.