Quantitative Evaluation of Facial Expressions and Movements of Persons 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
jnaru@koto.kpu-m.ac.jp

Abstract

The video is analyzed using image processing and the feature parameters of facial expressions and movements, which are extracted in the mouth area. The feature parameter for expressing facial expressions is defined as the average of facial expression intensity. That for expressing movements of a person is defined as the average of absolute value of vertical coordinate for the center of gravity of mouth area in the relative coordinate system. The experimental result shows the usefulness of the proposed method.

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. Because of the growing trend of aging, the number of older people with dementia and/or depression living in rural area is increasing very rapidly. Due to mismatch between the number of patients and health care professionals there, it is difficult to provide psychological assessment and support for the patients living there.

We have been developing a method for analyzing facial expressions of a person while speaking with another person using a video phone to improve the QOL of elderly people living in care facility, or at home.[1-3] In the present study, we have developed a method for analyzing facial expressions and movements of a person while speaking with another person, using our reported method[2], the standardization of size of face to be analyzed, and the newly proposed feature parameters on facial expressions and movements of a person.

2. Proposed Method

2.1. *System overview and outline of the method*

The platform is composed of Skype[4] for video phone. In addition to record the audio and video dialogue, Netralia Pty Ltd's VodBurner[5] and Tapur[6] are introduced. The talks are recorded for the analysis of facial expressions and movements of a person. The recorded data are analyzed using image processing software, Open Source Computer Vision Library for real-time computer vision developed by Intel (Open CV)[7], the standardization of

size of face to be analyzed, and the newly proposed feature parameter on facial expressions and movements of a person described in the following subsections. The Y component obtained from each frame in the dynamic image is used for analyzing facial expressions and movements of a person. The proposed method consists of (1) Standardization of lower part of face-area in size, (2) extraction of mouth area, (3) measurement of facial expression intensity, (4) judgment of utterance, (5) calculation of feature parameter on facial expression strength, and (6) calculation of feature parameter on movements of a person. In the following subsections, these six are explained in detail.

2.2. *Standardization of lower part of face-area in size*

First, a face-area obtained from each frame in the dynamic image is extracted using the classifier for a front-view face included in OpenCV. In the classifier, Haar-like feature parameter and Adaboost algorithm for learning are used.[8] It can be assumed that the distance between a subject and the camera is almost always kept during conversation using Skype. The face of the frame where the face-area extracted using OpenCV has the minimum size among those for a period in the dynamic image is assumed to be the most likely front-view among those in the period,[1] and the frame is selected and used. In the present study, we set 1/3 seconds as the period. Then, a low part of face-area extracted by the above method is standardized in length and width for extracting a mouth area. This standardization in size is performed with the aim of not only improving the performance of extracting a mouth area using OpenCV but also normalizing the feature parameter generation using 2D-DCT (Discrete Cosine Transform) there. Under the circumstance that the size of face in a dynamic image cannot be kept constant every time, this standardization in size is indispensable for reliably measuring the feature parameters described in the sections 2.6 and 2.7.

2.3. *Extraction of mouth area*

Next, using OpenCV, the mouth area is extracted as a rectangle shape. The mouth area is selected because the difference between the facial expressions of neutral and happy remarkably appears there. Fig. 1 shows an example of face image, a lower part of face image before and after standardization of size, and mouth-area image extracted.

Fig. 1. Face image (upper), lower part of face image before (lower & left) and after (lower & center) standardization of size, and mouth-area image extracted (lower & right).

2.4. *Measurement of facial expression intensity*

For the Y component of the frame selected by the processing described above, the feature vector of facial expression is extracted in the mouth area with use of 2D-DCT performed for each domain having 8×8 pixels.

We select 15 low-frequency components of the 2D-DCT coefficients, except for a direct current component, as the feature parameters for expressing facial expression.[2] Then, we obtain the mean of the absolute value for each 2D-DCT coefficient component in the area of mouth.[2] Therefore, we obtain 15 values as the elements of the feature vector. The facial expression intensity, defined as the norm of the difference vector between the feature vector of the neutral facial expression and that of the observed expression, can be used for analyzing a change of facial expression.[2]

2.5. *Judgment of utterance*

The sound data are smoothed and sampled to erase noise. Then, all sampled data that fall within $\left[\bar{x}_s - 14\sigma_s, \bar{x}_s + 14\sigma_s\right]$, where $\bar{x}_s$ and σ_s respectively express the average and the standard deviation of the sound data value for one second under the condition of no utterance, are considered to be the range of no utterance. When at least one sampled datum has a value outside $\left[\bar{x}_s - 14\sigma_s, \bar{x}_s + 14\sigma_s\right]$, our system judges that the sound data contain an utterance after erasing the noise.

2.6. *Feature parameter on facial expression strength*

In diagnosing a patient having dementia and/or depression, it might be useful for health care professionals to evaluate the strength of facial expression using a simple measure. Moreover, it might be more advantageous for a diagnosis of dementia and/or depression to separately evaluate the strength of facial expression as a speaker and a listener. Therefore, we measure as the feature parameter of facial expression strength the average of facial expression intensity in the four cases of (1) both subjects A and B speak, (2) subject A speaks and subject B does not speak, (3) subject A does not speak and subject B speaks, and (4) both subjects A and B do not speak, using the method for judging an utterance described in the section 2.5.

2.7. *Feature parameter on movements of a person*

Head movements such as a nodding during conversation might suggest a mental state and/or recognition ability of a patient having dementia and/or depression. Therefore, we measure as the feature parameter of movements of a person the average of absolute value of vertical coordinate for the center of gravity of mouth area in the relative coordinate system in the four cases described in the section 2.6. The relative coordinate system is defined using the mouth area extracted at the starting point for measuring the feature parameter on movements of a person. At the starting point, the height of the mouth area is set to be one and the vertical coordinate for the center of gravity of the mouth area is set to be 0 in the relative coordinate system.

3. Experiment

3.1. *Condition*

Two males (subject A in his 50s and subject B in his 20s) participated in the experiment where they took conversation for about 70 seconds using Skype. The videos saved using VodBurner were transformed into AVI files, and the audios saved using Tapur were transformed into WAV files. The AVI files were used for measuring feature parameters on facial expressions and movements of a subject. The WAVE files were used for judgment of utterance. The size of image frame was 720×480 pixels, and the size of standardized lower part of face image was set to be 240×96 pixels.

3.2. *Results and discussion*

Fig. 2 shows mouth-area images at starting point. Facial expression intensity changes of subjects A and B during their conversation (Fig. 3) and changes of coordinates of center of gravity on mouth area (Fig. 4) are shown respectively. In Fig. 5, face-images are shown at the characteristic timing positions. Feature parameters on facial expressions and movements of subjects are shown in Table.1. The definitions of these parameters are described in the sections 2.6 and 2.7.

As shown in Figs. 3 and 5, facial expression intensity was very sensitive for facial expression change. Both of subjects A and B did not move vertically so much during conversation (Fig. 5 and Table 1). The number of mouth-area images extracted using OpenCV was increased for subject A from 196 to 197 by performing the size

Fig. 2. Mouth-area images of subjects A (left two) and B (right two) at starting point with (left side) and without (right side) size standardization before extracting mouth area.

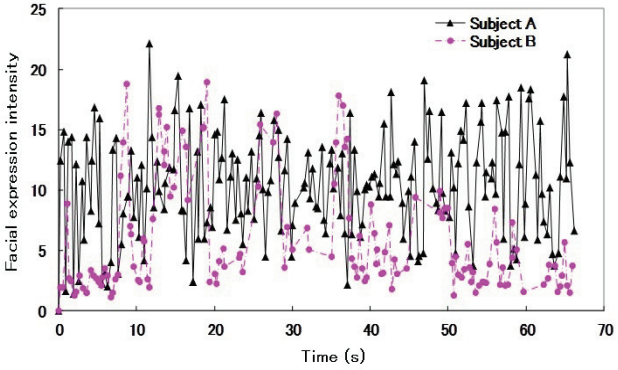

Fig. 3. Facial expression intensity change of subjects A and B during the conversation between these two subjects with size standardization before extracting mouth area.

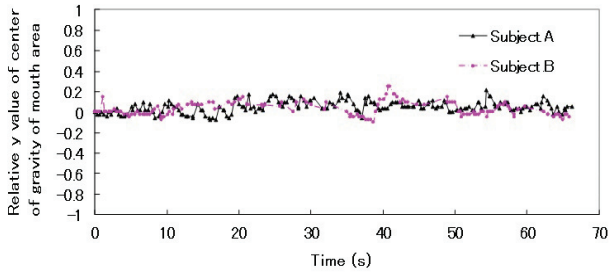

Fig. 4. Changes of coordinates of center of gravity on mouth area with size standardization before extracting mouth area.

Fig. 5. Face-images at characteristic timing positions on facial expression intensity value of subject A; upper: 0 (starting point), lower & left : maximum, lower & right : minimum except that at starting point.

Table 1. Feature parameters on facial expressions and movements.

(1) With size standardization before extracting mouth area

Subject	Utterance		Feature parameter	
	A	B	Facial expression	Movement of person
	without	without	11.41	0.05
A	with	without	9.47	0.08
	without	with	10.50	0.05
	B	A		
	without	without	4.53	0.05
B	with	without	12.51	0.07
	without	with	3.14	0.06

(2) Without size standardization before extracting mouth area

Subject	Utterance		Feature parameter	
	A	B	Facial expression	Movement of person
	without	without	8.55	0.03
A	with	without	8.50	0.04
	without	with	10.06	0.05
	B	A		
	without	without	5.48	0.04
B	with	without	10.88	0.05
	without	with	5.18	0.03

standardization before extracting mouth area, while it was increased for subject B from 27 to 138 by the standardization. Though mouth-area images were influenced by size standardization before extracting mouth area (Fig. 2), the feature-parameter values of both facial expressions and movements of subjects were not influenced so much by size standardization before extracting mouth area (Table 1). The value of feature parameter of facial expression was relatively high for subject A in the all three cases, while it was relatively high for subject B in the only case that subject B spoke and subject A did not speak (Table 1). Because the period in which both subjects A and B spoke was very short, the data were not described in Table 1.

4. Conclusion

The video is analyzed using image processing and the newly proposed feature parameters of facial expressions and movements. The experimental result shows the usefulness of the proposed method. We will develop the method for estimating a mental state and/or recognition ability of a patient using the proposed method.

Acknowledgment

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

References

1. T. Asada, Y. Yoshitomi, A. Tsuji, R. Kato, M. Tabuse, N. Kuwahara, and J. Narumoto, Method of facial expression analysis of person while using video phone, *in Proc. of Human Interface Symposium 2013* (Japan, Tokyo, 2013), pp.493-496.
2. T. Asada, Y. Yoshitomi, A. Tsuji, R. Kato, M. Tabuse, N. Kuwahara, and J. Narumoto, Facial expression analysis while using video phone, *in Proc. of Int. Conf. on Artif. Life and Robotics* (Japan, Oita, 2014), pp.230-234.
3. J. Narumoto, N. Kuwahara, Y. Yoshitomi, T. Asada, Y. Kato, H. Kamimura, K. Fukui, Development of support system for patients with dementia through teleconference system, *in Proc. of 4th World Conf. of Asian Psychiatry (*Thailand, Bangkok, 2014*)*, pp.1-4.
4. Skype Web page, http://www.skype.com/ Accessed 5 November 2013.
5. VodBurner Web page, http://www.vodburner.com/ Accessed 11 July 2014.
6. Tapur Web page, http://www.tapur.com/jp/ Accessed 18 December 2014.
7. OpenCV Web page, http://opencv.willowgarage.com/ Accessed 11 July 2014.
8. P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, *in Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (USA, Kauai, 2001), Vol.1, pp.511-518.