ATLANTIS PRESS Journal of Robotics, Networking and Artificial Life, Vol. 4, No. 2 (September 2017) 114–117

A System for Analyzing Facial Expression and Verbal Response of a Person While Answering Interview Questions on Video

Taro Asada, Yasunari Yoshitomi, 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@mei.kpu.ac.jp, {yoshitomi, tabuse}@kpu.ac.jp
http://www2.kpu.ac.jp/ningen/infsys/English_index.html

Abstract

We have developed a system for analyzing facial expressions of a person while answering interview questions on a video. A video capturing the answerer is analyzed by OpenCV and the feature parameter for an eye area in addition to the mouth area focused in our reported research. Moreover, the time to utterance of the person answering just after an interview question and the fundamental frequencies of his or her voice are measured for analyzing his or her mental state.

Keywords: Facial expression analysis, Movement analysis, Mouth area, Eye area, Interview on video, OpenCV.

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 has been increasing very rapidly. 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] However, in our previously reported system[1,2] for analyzing the facial expression of a person during a conversation with another person, facial expression intensity was assumed to be affected by (1) the conversation topic, (2) the partner, and (3) the facial expression of the partner. In the present study, (1), (2), and (3) are fixed by using an interview video. We developed a system for analyzing the facial expressions of a person while answering the interview questions on video. The video capturing the "answerer" in an interview is analyzed by image processing software (OpenCV[3]) and the previously proposed feature parameter (facial expression intensity), which is measured for the eye area in addition to the mouth area. The mouth area was the focus of our previously reported research. Moreover, the time to utterance of the answerer just after an interview question and the fundamental frequencies of his or her voice are measured for analyzing his or her mental state.

2. Proposed System and Method

2.1. *System overview and outline of the method*

We constructed three modules in our system based on our reported research.[4] The first is a module for replaying an interviewer's video for the answerer, the second is a module for recording the answerer's

response on video, and the third is a module for processing the recorded video of the answerer. The first module is a pseudo-interview. In the recorded data, the sizes of the faces are standardized, and the data are analyzed by using OpenCV for some feature parameters, as outlined below.

The Y component obtained from each frame in the dynamic image is used for analyzing the facial expressions. The proposed method is as follows: (1) the size of the lower part and the upper part of the face are standardized; (2) the mouth area in the lower part and the eye area in the upper part of the face are extracted; (3) the facial expression intensities for these two areas are measured; (4) the reference frame is selected; (5) the best positions for the mouth-part and the eye areas in the frame are determined; (6) the time to utterance of the answerer just after a given interview question is measured for analyzing the mental state of the answerer; (7) the feature parameters for the facial expression strengths of these two areas are calculated; and (8) the fundamental frequencies for voices of the answerer are measured by a conventional method. In the following subsections, (2) and (3) are explained in detail. For details on (1), (4), and the judgment of utterance in (6), see Refs. 1 and 2. For details on (5) of the mouth area, see Ref. 2. For details on (7) of the mouth area, see Ref. 1. The best positions for the eye areas in the frame are determined in a way similar to that for the mouth area. In addition, the feature parameters for facial expression strengths of the eye areas are calculated in a way similar to that for the mouth area.

2.2. *Extraction of the mouth-part and eye area*

Next, by using OpenCV, the mouth-part and eye areas are extracted as rectangular shapes. An example of a face image and the extracted images of the mouth-part and eye areas is shown in Fig. 1.

2.3. *Measurement of facial expression intensity*

For the Y component of the selected frame, the feature vector for facial expression was extracted for the mouth-part and eye areas. The extraction was performed by using a discrete cosine transform (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, but excluded the direct current component. This selection of 2D-DCT coefficients is popular among researchers in facial expression recognition.[5] Next, we obtained the mean of the absolute value of each of these components in the areas of the mouth-part and the eye-part. In total, we obtained 15 values as elements of the feature vector for the areas of the mouth-part and the eye-part. 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 expressionless face and that for an observed expression.[6] The candidate of facial expression intensity, defined as the norm of the difference vector between two feature vectors, was used for selection of the reference frame.[2]

Fig. 1. Whole-face image (left), extracted images of the eye area (upper right) and the mouth area (lower right).

3. Experiments

3.1. *Conditions*

The interviewer's video was made from a video recording a scene in which a psychiatrist interviewed a subject. Only the psychiatrist's voice remains in the interviewer's video. In the video, the psychiatrist asks the subject three fundamental questions generally used for a patient with potential depression. Based on a preliminary experiment for deciding the appropriate interval time, the interval time between two questions was edited by using video editing software. The experiment was performed in the following computational environment: the PC set in front of subject A was a Dell Precision T1600; CPU: Intel Xeon E31225 3.1 GHz; 8.0 GB memory; OS: Microsoft Windows 7 Professional. The development language was Microsoft Visual C++ 2008 Express Edition.

Two males (subject A in his 30s, subject B in his 20s) participated in the experiments. The subjects were interviewed by the interviewer video, which was about 40 seconds in length. As an initial condition in the experiment, the subjects were instructed to keep a

neutral facial expression without utterance for about five seconds just after starting the playback of the interview video. After the initial state of the neutral facial expression was terminated, the subjects were requested to intentionally respond with three types of emotions (Experiment 1: relaxed, Experiment 2: excited, Experiment 3: depressed). The visual and audio information of the subjects in the experiments were saved as AVI files having 640×480 pixels for each frame. The AVI files were used for measuring the feature parameters of the facial expressions, the time to utterance of the subject just after an interview question was given, and the fundamental frequencies of the subject.

3.2. *Results and discussion*

The facial expression intensity for the mouth area was more sensitive than that for the eye area (Figs. 2-5). The facial expression intensity for the mouth area became high during the response to a question. On the other hand, the facial expression intensity for the eye area fluctuated widely and was almost independent of the utterance. Note that the feature parameters for facial expressions for the mouth area tended to be greater in Experiment 2 than in both Experiments 1 and 3, while those for the eye area tended to be smaller in Experiment 1 than in both Experiments 2 and 3 (Table 1). The average of the fundamental frequencies of the voice of the subjects was greater in Experiment 2 than in both Experiments 1 and 3 (Table 2). The order of increasing time to utterance of the answerer just after an interview question was terminated was Experiments 2, 1, and 3 (Table 3). The experimental results described in this subsection show the usefulness of the proposed system.

Table 1. Feature parameters for facial expressions.

Exp.		1(relaxed)	2(excited)	3(depressed)
Subject A	Mouth-part	1.5	1.8	1.7
	Eye-part	2.4	3.6	4.4
Subject B	Mouth-part	2.2	3.5	1.4
	Eye-part	3.4	4.3	3.6

Table 2. Average of fundamental frequencies of voice of the subjects for each intentional emotion in Experiments 1-3.

Exp.	1(relaxed)	2(excited)	3(depressed)
Subject A	127.97	144.61	107.03
Subject B	93.33	137.38	99.55

(Hz)

Table 3. Time to utterance of the answerer just after an interview question was terminated.

	Exp.	Q.1	Q.2	Q.3	Ave.
Subject A	1(relaxed)	2.3	1.3	0.8	1.4
	2(excited)	0.6	1.0	0.0	0.5
	3(depressed)	2.5	2.6	2.1	2.4
Subject B	1(relaxed)	1.7	1.5	1.1	1.4
	2(excited)	0.8	0.7	0.3	0.6
	3(depressed)	1.7	2.0	1.8	1.8

(s)

4. Conclusion

We developed a system for analyzing facial expressions of a person while answering interview questions in a video. Moreover, the time to utterance of the answerer just after an interview question and the fundamental frequencies of his or her voice were measured in order to analyze his or her mental state. The experimental results show the usefulness of the proposed system.

Acknowledgements

The authors would like to thank Professor J. Narumoto of Kyoto Prefectural University of Medicine for his valuable support and helpful advice in the course of this research. We would also like to thank Mr. K. Nishimura, a student of the Graduate School of Kyoto Prefectural University, for his cooperation in the experiments. 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, A system for facial expression analysis of a person while using video phone, *J. Robotics, Networking and Artif. Life* **3**(1) (2016) 37-40.
3. Open CV. http://opencv.org/ Accessed 18 November 2016.
4. T. Asada, Y. Yoshitomi, M. Tabuse and J. Narumoto, A system for facial expression analysis of a person while answering interview questions on video (in Japanese), in *Proc. of Human Interface Symposium 2016* (Japan, Tokyo, 2016), pp.679-682.
5. T. Sakaguchi and S. Morishige, Real-time facial expression recognition based on the 2-dimensional DCT, *Trans IEICE* ***J80-D-II*** (6) (1997) 1547-1554.
6. T.Asada, Y.Yoshitomi, A.Tsuji, R.Kato, M.Tabuse, N.Kuwahara, and J.Narumoto, Facial expression analysis while using video phone, *J. Robotics, Networking and Artif. Life*, **2**(4) (2016) 258-262.

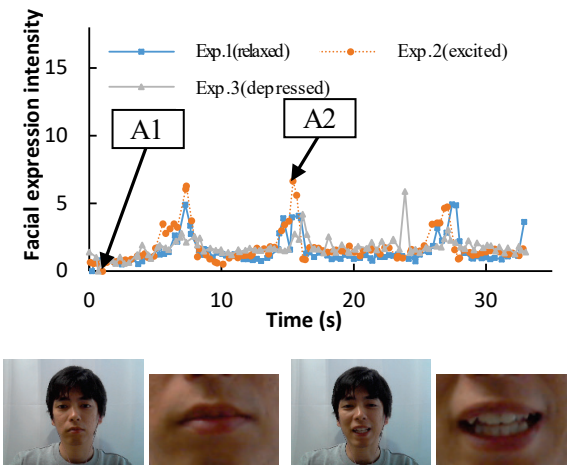
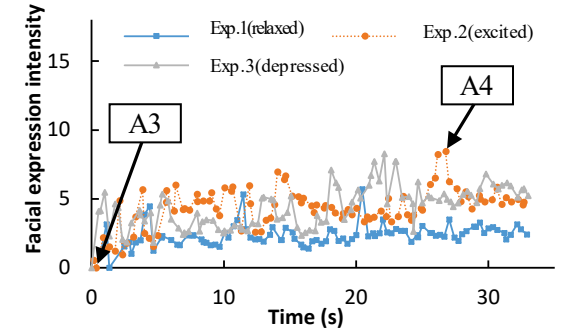


Fig. 2. Changes in facial expression intensity of mouth area for subject A (upper graph). Whole-face images and mouth images are shown for two moments during Experiment 2 (A1 and A2), as indicated on the graph (lower images).



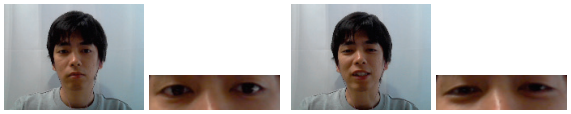
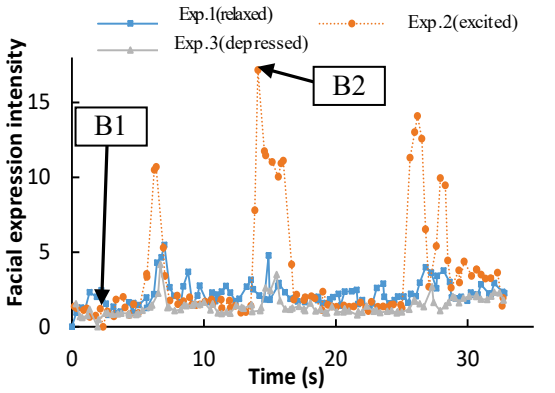


Fig. 3. Changes in facial expression intensity of eye area for subject A (upper graph). Whole-face images and eye-part images are shown for two moments during Experiment 2 (A3 and A4), as indicated on the graph (lower images).



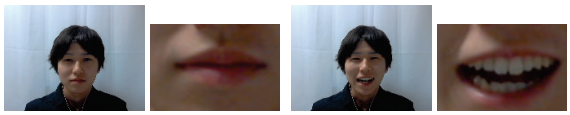
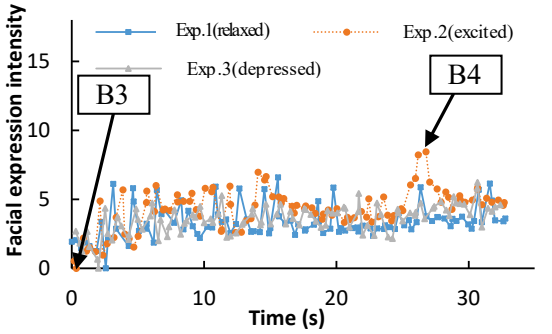


Fig. 4. Changes in facial expression intensity of mouth area for subject B (upper graph). Whole-face images and mouth images are shown for two moments during Experiment 2 (B1 and B2), as indicated on the graph (lower images).



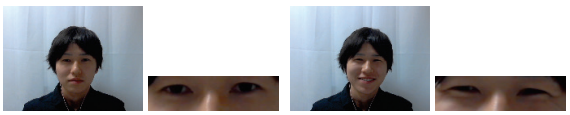


Fig. 5. Changes in facial expression intensity of eye area for subject B (upper graph). Whole-face images and eye-part images are shown for two moments during Experiment 2 (B3 and B4), as indicated on the graph (lower images).