Facial expression analysis using motion capture and eye tracking

Masakazu MURATA, Hajime HARADA, Takashi X. FUJISAWA, and Noriko NAGATA

Kwansei Gakuin University, nagata@kwansei.ac.jp

Abstract Humans are well known to use non-verbal information as well as verbal information in communication. Non-verbal information is expected to play an important role relatively in complex communication including agreement and apology. We propose an analysis method for vague facial expression using motion capture and eye tracking simultaneously. Eye directions and facial movements of a subject while two kinds of face-to-face communication, 'free' conversation and 'compulsory' conversation, are recorded and analyzed by Self-organizing map (SOM). The results of the analysis show that four kinds of expression of affirmation, negation, interrogation and thought, have their own features in the facial motions and the eye directions.

1 Introduction

The realistic expression of the human body using computer graphics (CG) is desirable in various fields such as industries providing interactive content and human interactions, in addition to the movie, television, and game industries. In particular, facial expression has been considered as one of the critical issues since the initial development of CG. Several researches have been conducted on these issues [2] in the past.

A typical example of previous developments is the concept of AU (action unit) and FACS (Facial Action Coding System) that was developed by Ekman, et al. [3]. AU is defined as "the minimum unit of facial expression motion identifiable by human visual function" and is not included in the anatomical characteristics. FACS is a method used for expressing facial expressions in combination with AUs. Since it is possible to express various facial expressions by a small number of parameter operations, FACS continues to be used as a means for expression synthesis. However, irrespective of these advantages, since the AU has been categorized and defined subjectively by various facial expressions, the synthesized facial expressions sometimes give artificial and unnatural impressions. This shows that the synthetic method using AUs has limitations in expressing certain realities.

Humans commonly use nonverbal information as well as verbal information in communication. Nonverbal information is expected to play a relatively important role in complex forms of communication such as agreements and apologies. We propose an analysis method for vague facial expressions that simultaneously uses motion capture and eye tracking. In this research, we focused on the facial expressions and eye directions obtained from nonverbal information and performed simultaneous measurements during face-to-face communications. A motion capture system was used for the extraction of a specific expression from the facial expressions, while an eye tracking system was used for the extraction of the eye line. The objective of this research was to define the correlation between the two. The subjects were asked to engage in two types of face-to-face communication: (1) "free" conversation and (2) "compulsory" conversation. During the communications, their facial motions and eye directions were measured. The measurement data of the facial motion and eye direction were cut off before and after the appearance of expressions to study the characteristics of 4 expression types. Lastly, the

facial expressions were categorized using self-organizing maps (SOM), a teacherless learning model, to indicate the efficiency of the simultaneous measurement of the facial motions and eye directions.

2 Related Researches

Only a few analyses have been reported on the correlation between the facial expression and the eye line [5], created an eyeball movement model based on the result of measurements taken by measuring the facial expressions and eye directions during face-to-face communication between two people; they then produced a realistic facial animation by synthesizing it with the motion of facial expressions. Their model focused on the characteristics of saccade (rapid movement of the eye), a high-speed eye movement, and their results agreed with the findings regarding eyeball movement that were presented in previous researches in the fields of psychology and cognitive science. However, their research focused on building an actual eyeball movement model that strongly corresponds to the measurement result; therefore, the facial model motion measurement is rather supplementary as it is only made to improve the precision of the model. As such, the facial expression and eye direction were not simultaneously measured. Further, since a series of communications lasting a few minutes each were used for the analysis of the measurement result, the research did not find the correlation between the expression motion of a specific expression and the eye direction. Since few researches have been conducted from the viewpoint of clarifying the characteristics of the facial motion and eye direction at the characteristic manifestation of an expression, this research was conducted with the aim of clarifying the correlation between the two.

3 Simultaneous Measurement of Facial Motions and Eye Lines

3.1 Measurement Method

In order to simultaneously measure the facial motions and eye lines, an optical motion capture system and an eye tracking system were used in combination. The motion capture system was used to measure the facial motion and head position and the eye capture system was used to measure the eye direction. The measurement methods of each system are described below.

3.1.1 Motion Capture of Facial Expressions

(1) Optical marker placement: As shown in Figure 1, optical markers are placed at 51 points on the face of the subject and 3 points on the head. The characteristic points for placement were set by referring to the FP (future point) in the FDP (facial definition parameter) [3].

(2) Camera Position: Optical motion capture system (manufactured by Motion Analysis) infrared cameras were placed on the left and right sides to the front of the subject; there were 4 units on each side for a total of 8 units, as shown in Figure 2. The cameras were set in an arc shape to avoid the interference of infrared rays and to enable the recognition of all facial markers. Further, the cameras were not placed immediately to the front of the subject to avoid face-to-face communication interference.

(3) Head movement correction: In order to cancel the movement of the head, facial markers were corrected by the motion data for 3 points of the head obtained by motion capture.

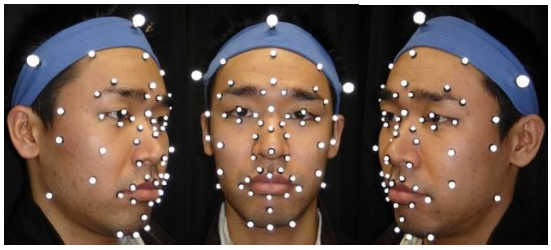


Fig. 1 Optical marker placement.

Fig. 2 Infrared camera placement.

Fig. 3 Eye picture.

3.1.2 Measurement of Eye Direction

The eye tracking system determines the eye direction based on the eye movement and the calibration of gazing points in the visual plane. In this research, a video camera was placed to the left of the subject's head to capture images of the face of the person with whom face-to-face communication was being conducted. Using this picture as the picture in the visual plane, the gazing points in the visual plane, i.e., the eye direction, were calculated.

3.2Facetoface Communication Experiment

3.2.1 Subject

The subjects were 3 males aged 23 to 24, and the persons with whom face-to-face communication was conducted were 4 males aged 21 to 31. Additionally, persons communicating with the subjects were acquaintances belonging to the same community.

3.2.2 Details of Communication

In this research, the face-to-face communications conducted for the purpose of measurements comprised the following 2 patterns.

(1) Free conversation: Communications were conducted without specifying the topic. The measurement time was approximately 3 min.

(2) Compulsory conversation: Conversations that encouraged the manifestation of 9 types of expressions such as affirmation, negation, interrogation, confirmation, and thought during free conversation are defined as compulsory conversation. For these 9 types of conversations, two topics were prepared for each type to enable the random interchange of 18 short communications. Each conversation comprised one question and one response and 1 conversation lasted approximately 5s. The abovementioned communications with 2 patterns were conducted for 2 people.

3.3 Analysis Data

Among the 9 types of conversations, both free conversation and compulsory conversation were measured in good balance. Analyses were performed on 4 expressions of affirmation, negation, interrogation, and thought. The expression manifestation for this research was judged by typical utterances (if affirmative, "yes" and "uh-huh"; if negative, "no" and "uhn-uhn"; if interrogative, "is it ~?" and "you mean ~?"; if thoughts, "well" and "um ~ "). Some examples of conversation are given below.

(1) Examples of free conversation

- "You have been hired, right?" "Yeah" (Affirmation)
- "I am not very good at talking ----," "No, you're good at it" (Negation)

(2) Examples of compulsory conversation

- "Um----" -> "What is it?" (Interrogation)
- "What did you have for supper last night?" -> "Well, um --- I had __" (Thought)

With regard to each expression, 30 frames before the starting frame of expression manifestation and 90 frames after the manifestation, i.e., a total of 120 frames, were extracted. The starting frame of expression manifestation was judged as the point of time at which the utterance applicable to each expression began. In this research, eye blink information was not used. Therefore, data not containing eye blinks was used for extracting the samples. Since the sampling rates for both motion capture and eye tracking are 60 Hz, the analysis lasted for 2s.

The numbers of data for each expression were: affirmation 5 (free:1, compulsory:4), negation 6 (free:2, compulsory:4), interrogation 8 (free:4, compulsory:4), and thought 8 (free:4, compulsory:4), totaling 27 data sets.

3.4 Analysis Results

3.4.1 Eye Direction

Eye directions measured with regard to 4 expressions, namely, affirmation, negation, interrogation, and thought, are plotted on the graph (indicated by changing the shapes and colors of the markers for each of the 3 subjects). The outside

frame indicated by the thick line represents the "eye field" that indicates the maximum visual field of the subject while the inside frame representing the "face area" indicates the range where the face of the other person exists.

Since no large difference was observed in any communication other than negation (affirmation, interrogation, and thoughts), all conversation data were totalized and plotted in one graph. For the expression of "negation," the differences that were clearly observed are shown in the graph (Fig. 5).

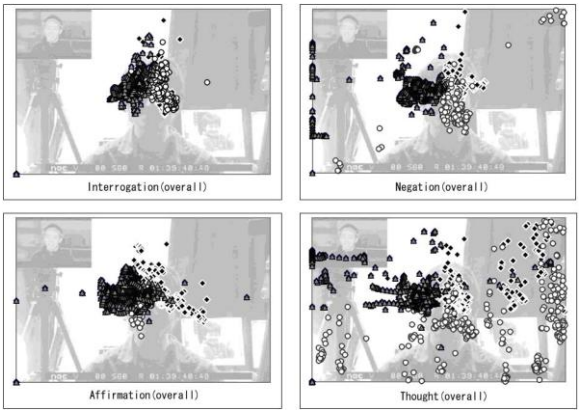


Fig. 4 Eye line distribution of expression (free conversation/ compulsory conversation).

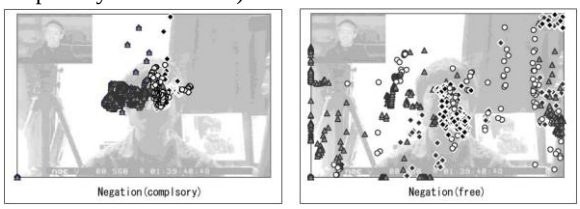


Fig. 5 Eye line distribution of negation expression (b1: free conversation, b2: compulsory conversation).

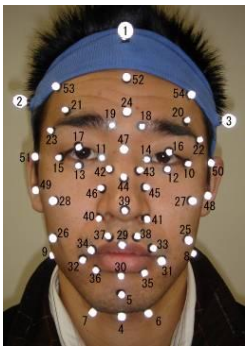


Fig. 6 Allocation of marker numbers.

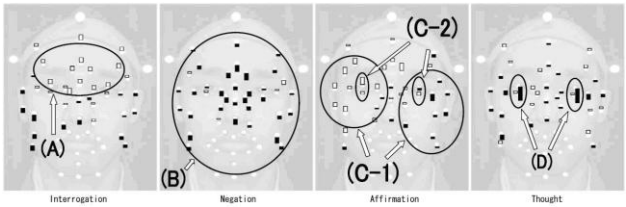


Fig. 7 z-Coordinate variation value of expression.

3.4.2 Facial Motion

Variations were calculated from the coordinate values (x, y, and z) of each marker between the frames at expression manifestation and the average value was determined for each expression from all data of 3 subjects. The marker positions are indicated by numbers. The allocations are shown in Figure 6. For 2 frames used for calculating the variation, the 30th and 31st frames were selected at the start of the expression manifestation. These frames were selected because they were the closest frames to the start of expression manifestation having no influence of eye blinking on the data. The coordinates facing the subject's face are the x-coordinate along the longitudinal direction, y-coordinate along the horizontal direction, and z-coordinate along the vertical direction. In the identification of 4 types of expressions, a characteristic change was observed along the z-axis (vertical direction). The result is shown in Figure 7. The white and black bars in the drawing indicate the positive and negative directions, respectively. The length of the bar indicates the variation.

3.5 Discussion

3.5.1 Eye Direction

As shown in Figure 3, a tendency of averting eyes from the face area was consistently observed in the expression of "thought" for all subjects. On the contrary, in the expressions of "affirmation" and "negation," a tendency of focusing the eyes on the face area was identified. These results conform to the experiences in daily life from a sensual point of view, and eye-gaze avoidance behavior when thinking has been observed in previous studies [6].

For the expression of "negation," the results were different between free conversation and compulsory conversation (Fig. 5). First, during free conversation for "negation," the same eye-gaze avoidance behavior was observed as in the case of the expression of "thought." On the contrary, in compulsory conversation, the eye line was focused on the face area as in the cases of the expressions of "affirmation" and "interrogation." The reason for the difference in these results is that in free conversation, a clear and predicable negation did not

exist as in the case of compulsory conversation. Although the expression type of utterance was "negation," when we consider the context, we observe that it has a strong meaning of a temporary response for taking time to consider what to speak next. In this sense, it is considered acceptable for the same result as "thought" to be obtained.

3.5.2 Facial Motion

With regard to facial motion, the markers on the upper eyelids of both eyes (16 and 17) exhibited remarkable values along the negative direction (Fig. 7A) for the expression of "thought." These can be considered as half-closing movement of the eyes, and can be identified from the fact that the values of the lower eyelids (markers 12 and 13) are both positive values. Next, "negation" can be considered as an expression that exhibited remarkable results. This expression is associated with the lowering of the eyes irrespective of the facial components since it mostly exhibits negative values (Fig. 7B).

In the expression of "affirmation," contrary to "negation," positive values were observed at the upper half of the face, i.e., in the inner end of eyes (markers 11 and 14) and upper eye lids (markers 16 and 17) and the inner end of eyebrows (markers 18 and 19) (Fig. 7A). It has been identified to have an increasing trend. In contrast, at the lower half of the face, a decreasing trend similar to that in the expression of "negation" was observed. As such, it can be said that this is an expression of giving consent that is made by slightly sticking out the face.

Finally, in the expression of "interrogation," left and right asymmetry in the expression behavior has been identified (Fig. 7C-1). In particular, when we focus on the behaviors of both the upper and lower eyelids of the left eye (makers 12 and 16) and the right eye (markers 13 and 17), the left eye appears to be closing but the right eye appears to be opening wide (Fig. 7C-2). It has been reported in many previous researches that there is a correlation between interrogation and hesitation that does not accompany a clear conclusion as in the cases of affirmation and negation, including insecure ambivalent emotions and the manifestation of asymmetrical facial expression. The fact that the same tendency has been identified during the "interrogation" expression manifestation in this research is very interesting in the sense that we were able to extract complicated expressions that are not included in the basic 6 expressions. The characteristics of eye directions and facial motions in each expression category are given in Table 1.

4 Categorization of Facial Expression by SOM

Using the eye direction and facial motion data, a self-organizing map (SOM) was created to categorize facial expressions. We discussion the extent to which the eye direction data contributes to the categorization of facial expressions.

4.1 Pre-processing

The characteristic amount for use in learning was selected. First, for the eye direction data, sections lasting 0.5 s obtained stably from all 2s data were selected and the eye coordinate movement value (x, y coordinates, 18-dim.) for every 0.05 s was used.

For the facial motion, correlations are determined on a total of 54 markers used in the measurement of facial motions. The coordinate values (x, y, z coordinates) of 19 pieces of markers with low correlation efficiency were used.

Learning was conducted using these data and the categorization result was visualized. Learning was conducted on the following 3 types of data sets. (a) Only eye direction data (18-dimension), (b) only facial motion data (57-dim.), and (c) 75-dim. data set combining eye direction and facial motion. For the map, 10 × 10 nodes were set and the initial positions was randomly assigned. The learning frequency was 30000 times.

Table 1. Expression and Various Characteristics of Eye Line and Facial Motion

	Eye Direction	Facial Motion
Affirmation	Face of the opposite person	Lifting of upper half, dropping down of lower half
Negation	Face of the opposite person	Dropping down
	Avoid	
Interrogation	Face of the opposite person	Left-right asymmetry
Thought	Avoid	Narrow the eyes

4.2 SOM Categorization Result

The result is shown in Figure 8. The thickness of the grid color indicates the distance between the grids. The distance decreases as the grid gets whiter.

4.3 Discussion

In (a), "thought" and "negation" are placed at the centre and "affirmation" and "interrogation" are placed at the periphery. As noted in Chapter 5, these are considered to reflect the movement of turning the eye direction toward the face of the other person and turning it away, according to personal emotions.

With regard to (b), we focus on "thought" and "negation" is positioned around it. This is considered to express the characteristics that these 2 expressions have of moving the face downwards. However, for (a) and (b), "affirmation" and "interrogation" were not adequately categorized.

On the contrary, (c) is observe to be categorized into 3 categories: "affirmation," "interrogation," and "thought" and "negation." In other words, it has been confirmed that by using 2 data sets of facial motions and eye directions, the precision of facial expression categorization can be improved. By increasing the data, further improvements in categorization precision can be expected.

5 Conclusion

In this paper, for the 4 main expressions observed in face-to-face communication, we simultaneously measured the facial motions and eye directions in order to consider the characteristics of each expression. Further, the facial expressions were categorized by using SOM and it has been proven that the categorization precision improves with the use of the data of both facial motions and eye directions. Further, the effectiveness of simultaneous measurements has been identified.

The 4 expressions addressed here are as clear as those in the basic 6 expressions addressed in previous researches, but depict more realistic and delicate human expressions. The results obtained show the possibilities of new developments in research concerning eye directions.

The issues for future development are to improve the credibility of the results identified in this research and to acquire more detailed results by increasing the number of subjects and obtaining more data. In such cases, it is desirable to reconsider the content of compulsory conversion and improve the measurement precision by fine tuning the eye tracking system. Further, in order to model the data obtained and create more realistic and situation-oriented facial expressions in the VR fields, the simultaneous expression of the eye lines and facial motions is also necessary. In the present research, subjects were selected from among acquaintances in the same community. However, in the future, we wish to further extend our study and seek interesting issues such as the extent to which personal likes and dislikes in human relationships affect the characteristics of eye directions and facial expressions.

References

[1] Waters, K.: "A muscle model for animating three-dimensional facial expression", Computer Graphics, 22(4), pp.17-24 (1987).

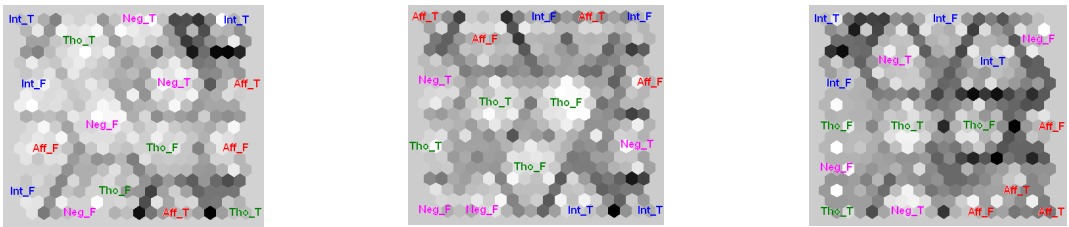
[2] Pandzic, I. S. &, Forchheimer, R.: "The MPEG-4 facial animation" John Wiley & Sons Inc (2002).

[3] Lee, S. P., Badler, J. B. & Badler, N. I.: "Eyes alive" , In Proc. of ACM SIGGRAPH (2002).

[4] Gale, A., Kingsley, E., Brookes, S. & Smith, D.: "Cortical arousal and social intimacy in the human female under different conditions of eye contact", Behavioral Processes, 3, pp. 271-275 (1978).

[5] Previc, F. H. & Murphy, S. J.: "Vertical eye movements during mental tasks: A reexamination and hypothesis", Perceptual & Motor skills, 84(3), pp. 835-847 (1997).

[6] Jaynes, J.: "The origin of consciousness in the breakdown of the bicameral mind", Mariner Books (1976).



(a) When only eye direction data set is used. (b) When only facial motion data set is used. (c) When the eye direction and facial data set is used.

Fig. 8 Categorization result by SOM.