Emotion Transmission System Using a Cellular Phone-Type Teleoperated Robot with a Mobile Projector

Yu Tsuruda, Maiya Hori, Hiroki Yoshimura, and Yoshio Iwai

Graduate School of Engineering, Tottori University
101 Minami 4-chome, Koyama-cho, Tottori 680-8550, Japan

Abstract. We propose an emotion transmission system using a cellular phone-type teleoperated robot with a mobile projector. Elfoid has a soft exterior that provides the look and feel of human skin and is designed to transmit a speaker's presence to their communication partner using a camera and microphone. To transmit the speaker's presence, Elfoid transmits not only the voice of the speaker but also their facial expression as captured by the camera. In this research, facial expressions are recognized by a machine learning technique. Elfoid cannot, however, physically display facial expressions because of its compactness and a lack of sufficiently small actuator motors. The recognized facial expressions are displayed using a mobile projector installed in Elfoid's head to convey emotions. We build a prototype system and experimentally evaluate its subjective usability.

1 Introduction

To communicate with people in remote locations, robots that have human appearance have been developed. Some studies have used humanoid robots for the transmission of human presence. In particular, teleoperated android robots such as Geminoid F and Geminoid HI-1 [1] have appearances similar to an actual person, and were intended to transfer the presence of actual people. These humanoid robots have high degrees of freedom and can transfer human presence. However, they are expensive and limited to a specific individual target. A robot called Telenoid R1 [2] was developed to reduce the number of actuators and costs. Telenoid is not limited to a specific individual target, and is designed to immediately appear and behave as a minimalistic human. A person can easily recognize Telenoid as human; it can be interpreted as male or female, and old or young. With this minimal design, Telenoid allows people to feel as if a distant acquaintance is next to them. Moreover, Telenoid's soft skin and child-like body size make it easy to hold. However, it is difficult to carry around in daily life.

For daily use, a communication medium that is smaller than Telenoid and uses mobile-phone communication technology is now under development. Like a cellular phone, Elfoid is easy to hold in the hand, as shown in Fig. 1. When we use such robots for communication, it is important to convey the facial expressions of a speaker to increase the modality of communication. If the speaker's facial movements are accurately regenerated via these robots, human presence can be

M. Kurosu (Ed.): Human-Computer Interaction, Part II, HCII 2014, LNCS 8511, pp. 707–714, 2014.

Fig. 1. Elfoid: cellular phone-type teleoperated android

conveyed. Elfoid has a camera within its body and the speaker's facial movements are estimated by conventional face-recognition approaches. However, it is difficult to generate the same expression in robots because a large number of actuators are required. Elfoid cannot perform facial expressions like a human face can because it has a compact design that cannot be intricately activated. That is, since Elfoid's design priority is portability, its modality of communication is less than Telenoid's. For this reason, it is necessary to convey emotions some other way.

2 Generation of Facial Expressions Using Elfoid

2.1 Elfoid: Cellular Phone-Type Teleoperated Android

Elfoid is used as a cellular phone for communication, as shown in Fig. 1. To convey the human presence, Elfoid has the following functions.

- Elfoid has a body that is easy to hold in a person's hand.
- Elfoid's design is recognizable at first glance to be human-like and can be interpreted equally as male or female, and old or young.
- Elfoid has a soft exterior that provides the feel of human skin.
- Elfoid is equipped with a camera and microphone.

Additionally, a mobile projector is mounted in Elfoid's head and facial expressions are generated by projecting images from within the head, as shown in Fig. 2.

2.2 Overview of the Total System

In this research, facial expressions are generated using an Elfoid's head-based mobile projector to convey emotions. Fig. 3 shows an overview of the total system.

First, individual facial images are captured using a camera mounted within Elfoid. Next, the facial region is detected in each captured image and feature points on the face are tracked using the Constrained Local Model (CLM) [3]. Facial expressions are recognized by a machine learning technique using the positions of the feature points. Finally, recognized facial expressions are reproduced using Elfoid's head-based mobile projector.

Fig. 2. Prototype system

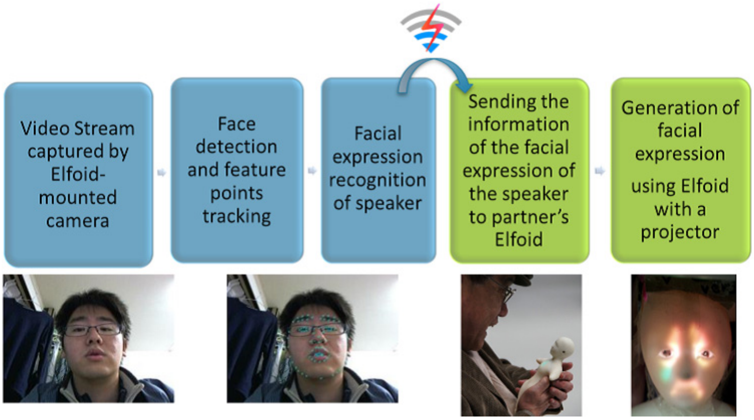


Fig. 3. Overview of the total system

2.3 Recognition of Facial Expressions

CLM fitting is the search for point distribution model parameters $\mathbf{p}$ that jointly minimize the misalignment error over all feature points. It is formulated as follows

$$\mathcal{Q}(\mathbf{p}) = \mathcal{R}(\mathbf{p}) + \sum_{i=1}^{n} \mathcal{D}_i(\mathbf{x}_i; \mathcal{I}), \tag{1}$$

where $\mathcal{R}$ is a regularization term and $\mathcal{D}_i$ denotes the measure of misalignment for the ith landmark at $\mathbf{x}_i$ in image $\mathcal{I}$. In the CLM framework, the objective is to create a shape model from the parameters $\mathbf{p}$. The misalignment term, $\mathcal{D}_i$, is estimated using the mean-shift technique. This method has low computational complexity and is robust to occlusion. The results of feature point tracking are shown in Fig. 4.

Fig. 4. Results of feature point tracking

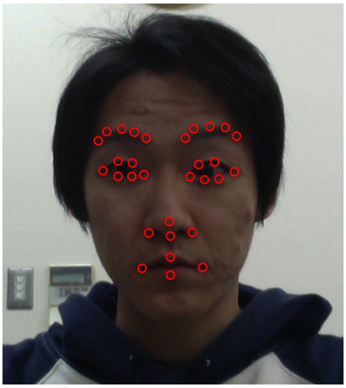


Fig. 5. Feature points used for classification

In this study, six facial expressions that correspond to universal emotions [4], happiness, fear, surprise, sadness, disgust, and anger, are classified using a hierarchical technique similar to [5]. The facial expressions are hierarchically classified by a Support Vector Machine (SVM). Each classifier is implemented beforehand using the estimated positions of feature points. The feature points used for classification are shown in Fig.5. This study is based on the theory that different expressions can be grouped into three categories [6,7] based on the parts of the face that contribute most toward the expression. These categories are shown in Fig. 6 At the first level, we use 31 feature points around the mouth, eyes, eyebrows and nose to discriminate the three expression categories: lip-based, lip-eye-based, and lip-eye-eyebrow-based. After grouping into three categories, each category is divided into two emotion classes. In the lip-based category, four feature points around the mouth are used for expressing happiness or sadness. In the lip-eye-based category, 16 feature points around the mouth and eyes are

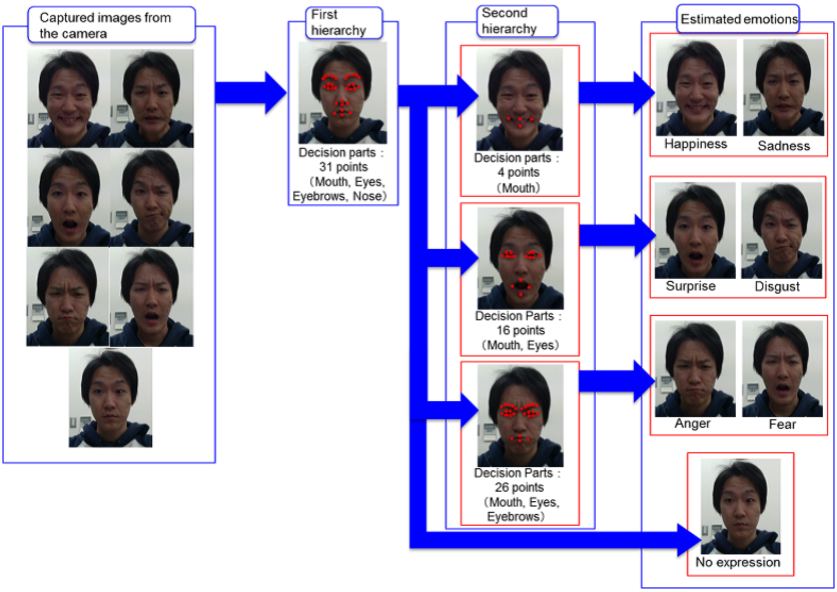


Fig. 6. Emotion estimation using a hierarchical technique

used for expressing surprise or disgust. In the lip-eye-eyebrow-based category, 26 feature points around the mouth and eyes and eyebrows are used for expressing anger or fear.

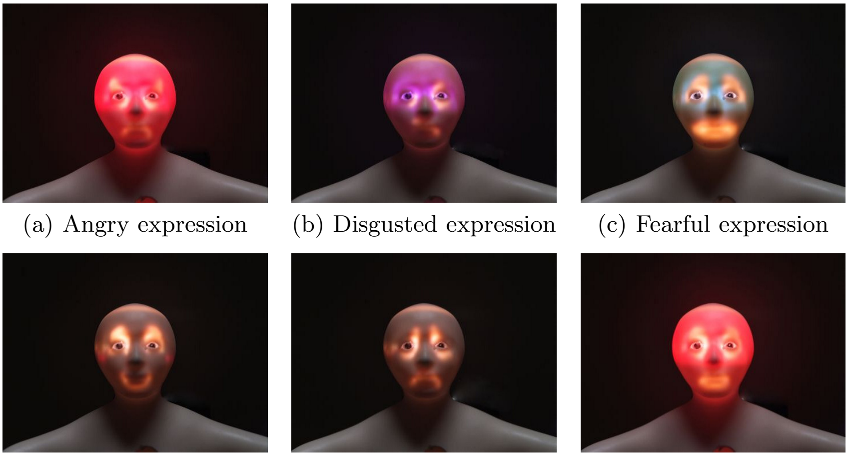
2.4 Generation of Facial Expressions with Elfoid Using Cartoon Techniques

Recognized facial expressions are reproduced using Elfoid's head-based mobile projector. To represent facial expressions, we generate stylized projection patterns using the results of emotion estimations. In this study, the projection patterns are stylized using cartoon techniques [8]. It is widely recognized that cartoons are very effective at expressing emotions and feelings. The movements around the mouth and eyebrows, for example, are exaggerated. The silhouette of face and shapes of eyes are varied by projection effects. Moreover, color stimuli that convey a particular emotion are added.

3 Experiment

3.1 Recognition Rate of Facial Expressions

We conducted an experiment to verify the recognition rate of facial expressions. As training data, we used a total of 8,000 images that consisted of 1,000 images for each facial expression and 2,000 images with no expression. To verify the rate



(a) Angry expression (b) Disgusted expression (c) Fearful expression
(d) Happy expression (e) Sad expression (f) Surprised expression

Fig. 7. Facial expressions generated by Elfoid

of the facial expression recognition, we tested 1,000 images for each expression that were different from the training data. Table 1 shows the facial expression recognition rate results. The results show that this estimation method can be applied to our communication system.

3.2 Subjective Evaluation of the Proposed System

In this experiment, representative face expressions were generated by Elfoid, as shown in Fig. 7. From the experiments also reported in [8] , it is shown that each emotion can be conveyed correctly. Additionally, to verify the validity of this system, we experimentally evaluated its subjective usability. Subjects had a conversation with a communication partner at a remote location using Elfoid. We used an Elfoid that does not project facial expressions as a comparison. We then gave the subjects questionnaires that asked about the satisfaction level

Table 1. Recognition rate for facial expressions (%)

answer \ estimation	happiness	sadness	surprise	disgust	anger	fear	no expression
happiness	**73.8**	3.0	0.0	0.0	23.2	0.0	0.0
sadness	1.4	**60.1**	0.0	0.0	33.9	0.0	4.6
surprise	0.0	0.0	**82.8**	0.0	17.2	0.0	0.0
disgust	0.0	0.0	0.0	**82.3**	17.7	0.0	0.0
anger	0.0	0.0	0.0	0.0	**100.0**	0.0	0.0
fear	0.0	0.0	1.9	0.0	0.2	**97.9**	0.0
no expression	0.0	0.0	0.0	0.0	10.4	0.0	**89.6**

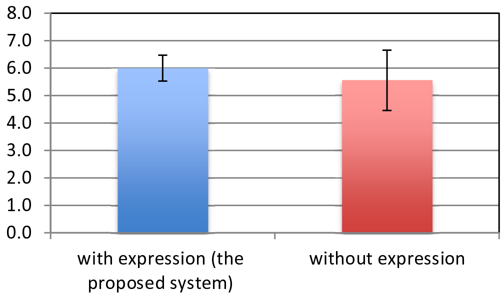


Fig. 8. User's satisfaction with the conversation

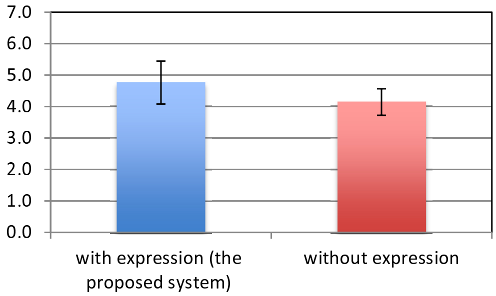


Fig. 9. User's impression of a conversational partner

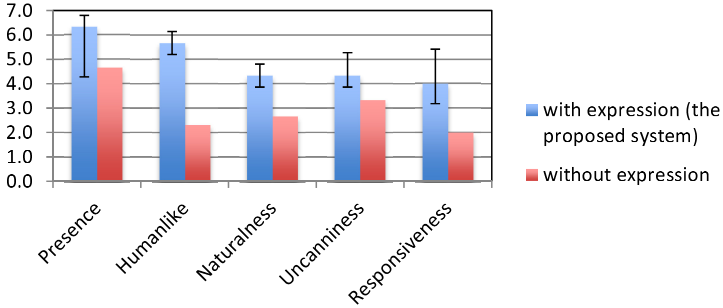


Fig. 10. User's impression of the interface

of the conversation, impression of a conversational partner, and impression of the interface. These items were based on the conventional method for assessing usability [9]. To ascertain the user's impression of the interface, presence, humanlike, naturalness, uncanniness, and responsiveness were investigated.

Figs. 8, 9, and 10 show the experimental results. Higher scores indicate a better impression. According to the results, the proposed system is effective for increasing conversation satisfaction level and the impression of a conversational partner. The evaluation regarding the impression of the interface is higher for the proposed system than the comparison system.

4 Conclusion

We propose an emotion transmission system using a cellular phone-type teleoperated robot with a mobile projector. In this research, facial expressions are recognized by a machine learning technique, and displayed using a mobile projector installed in Elfoid's head to convey emotions. In the experiments, we built a prototype system that generated facial expressions and evaluated the recognition rate of facial expressions and the subjective evaluations of usability. Given the results, we can conclude that the proposed system is effective for increasing conversation satisfaction level and the impression of a conversational partner.

Acknowledgment. This research was supported by the JST CREST (Core Research for Evolutional Science and Technology) research promotion program "Studies on cellphone-type teleoperated androids transmitting human presence."

References

1. Asano, C.B., Ogawa, K., Nishio, S., Ishiguro, H.: Exploring the uncanny valley with geminoid HI-1 in a real-world application. In: Proc. Int'l Conf. of Interfaces and Human Computer Interaction, pp. 121–128 (2010)
2. Ogawa, K., Nishio, S., Koda, K., Balistreri, G., Watanabe, T., Ishiguro, H.: Exploring the natural reaction of young and aged person with Telenoid in a real world. Jour. of Advanced Computational Intelligence and Intelligent Informatics 15(5), 592–597 (2011)
3. Saragih, J.M., Lucey, S., Cohn, J.F.: Deformable model fitting by regularized landmark mean-shift. Int'l Jour. of Computer Vision 91(2), 200–215 (2011)
4. Ekman, P., Friesen, W.V.: Unmasking the Face. Prentice-Hall, Inc. (1975)
5. Siddiqi, M.H., Lee, S., Lee, Y., Khan, A.M., Truc, P.T.H.: Hierarchical recognition scheme for human facial expression recognition system. Sensors 13, 16682–16713 (2013)
6. Nusseck, M., Cunningham, D.W., Wallraven, C., Bülthoff, H.H.: The contribution of different facial regions to the recognition of conversational expressions. Journal of Vision 8, 1–23 (2008)
7. Schmidt, K.L., Cohn, J.F.: Human facial expressions as adaptations: Evolutionary questions in facial expression resarch. American Journal of Physical Anthoropology 116, 3–24 (2001)
8. Tsuruda, Y., Hori, M., Yoshimura, H., Iwai, Y.: Generation of facial expression emphasized with cartoon techniques using a cellular-phone-type teleoperated robot with a mobile projector. In: Kurosu, M. (ed.) HCII/HCI 2013, Part V. LNCS, vol. 8008, pp. 391–400. Springer, Heidelberg (2013)
9. Sakamoto, D., Kanda, T., Ono, T., Hagita, N., Ishiguro, H.: Android as a telecommunication medium with a human-like presence. In: Proc. of the ACM/IEEE Int'l Conf. on Human-Robot Interaction, pp. 193–200 (2007)