Facial Mimicry Training Based on 3D Morphable Face Models

Oky Dicky Ardiansyah Prima, Hisayoshi Ito
Graduate School of Software and Information Science, Iwate Pref. Univ.
Takizawa, Japan
email: {prima, hito}@iwate-pu.ac.jp

Takahiro Tomizawa
Hitachi Ind. & Ctrl. Solutions
Yokohama, Japan
email: Takahiro.tomizawa.ax@hitachi.com

Takashi Imabuchi
Office of Regional Collaboration,
Iwate Pref. Univ.
Takizawa, Japan
email: t_ima@ipu-office.iwate-pu.ac.jp

***Abstract*— The recent techniques of automated facial expression recognition from facial images have achieved human perception levels. The application of this technology is expected not to be limited to facial expression analysis, but also to evaluate how well someone mimics another person's expression. Facial mimic training will help people improve their interpersonal communication and that, in turn, will improve their work performance. This study proposes a self-learning-based expression training system using a simple 3D Morphable Face Model (3DMM). The proposed system analyzes faces of a subject and a given picture of a person who the subject is mimicking. The 68 facial landmarks for both faces are detected automatically and are used to fit a 3DMM using a deformation transfer technique. Our experiment shows that the proposed system accurately measures the similarity of facial appearance between subjects and their corresponding mimic targets. Thus, the proposed system can be used as a facial mimicry training tool to improve social communication.**

Keywords-mimicry; expression training; emotion; image processing.

I. INTRODUCTION

Non-verbal (unspoken) communication plays an important role in providing additional information and cues over verbal communication. Facial expression is a type of non-verbal (spoken) communication that involves subtle signals of the larger communication process. For example, a smile, typically with the corners of the mouth turned up and the front teeth exposed, may indicate joy. Frowning, typically by turning down the corners of the mouth, forms an expression of disapproval.

While culture differences might cause differences in the absolute level of emotional intensity, the basic facial expressions such as happiness, surprise, sadness, fear, disgust, and anger are similar throughout the world [1]. The Facial Action Code System (FACS), which is based on the anatomical basis of facial movement, is a traditional measure to analyze facial expressions [2]. Individual facial muscle movements are encoded by FACS from slightly different instantaneous changes in facial appearance. Each Action Unit (AU) is described in the FACS manual.

Early attempts have been conducted to automate facial expressions using FACS. Bartlett et al. [3] applied computer image analysis to classify the basic elements that comprise complex facial movements. Their method classified six upper facial actions with 91% accuracy by combining three approaches: holistic spatial analysis, measurement of local facial features, and estimation of motion flow fields. However, this method was not fully automated, such as the initial facial alignment needs mouse clicks at the center of each eye. Tian et al. [4] developed an Automatic Face Analysis (AFA) system, which recognizes changes in facial expression into AUs. Initial detections of facial features, such as lips, eyes, brows and cheeks were done using template matching [5]. AFA has achieved around 96% recognition rates for upper and lower AUs, whether they occur alone or in combinations.

With the recent computer vision techniques, the conventional procedure to recognize facial expressions such as face detection, face alignment, facial feature extraction, and expression classification can be done in realtime. Affdex [6], one of the most widely used face analysis systems, provides a cross-platform realtime multi-face expression recognition. It uses Support Vector Machine (SVM) to train 10,000 manually coded facial images [7]. Affdex can achieve acceptable accuracy to detect facial expressions that are expressed externally on a face where certain parts of the face change significantly.

Automated facial expression recognition has been used in the development of humanoid robots to enable them to mimic human-like emotions [8]. Mimicking emotion is the act of imitating the facial expression of others and it is considered central for social interactions. The humanoid robot can be used as an experiment tool to construct a communication model of mimicry. A sophisticated model of mimicry will be useful to train people to improve social interactions through non-verbal communication [9].

Recently, 3D morphable models (3DMM) [10] have been widely used to create virtual faces in some software application programs, such as Augmented Reality (AR) and messaging apps. The advanced computer vision techniques have enabled to fit the model to the corresponding facial image. Moreover, it is possible to design virtual characters that can express different emotions such as compound emotions that are a mix of basic emotion expressions. Now, "Animoji"s exist, which are animated emoticons created by mirroring one's own facial expressions.

This study proposes a self-learning-based facial mimicry training system based on 3DMM to measure how close a person can mimic another person's facial expressions. The

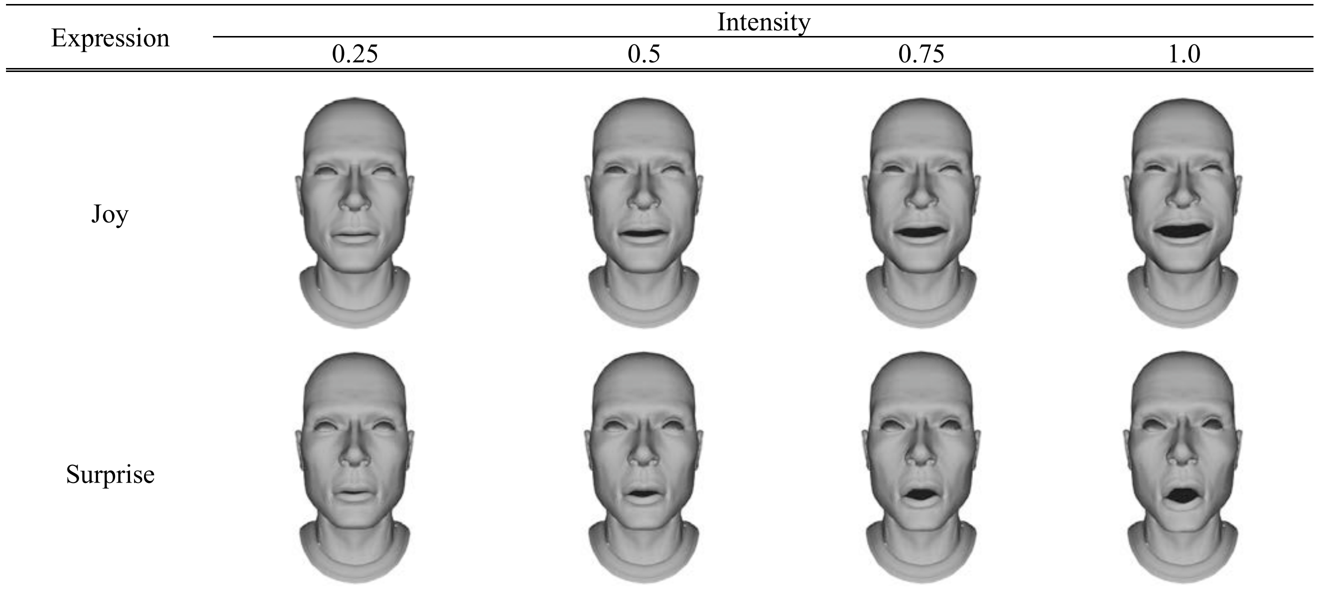

Figure 1. Some facial expressions generated using 3DMM.

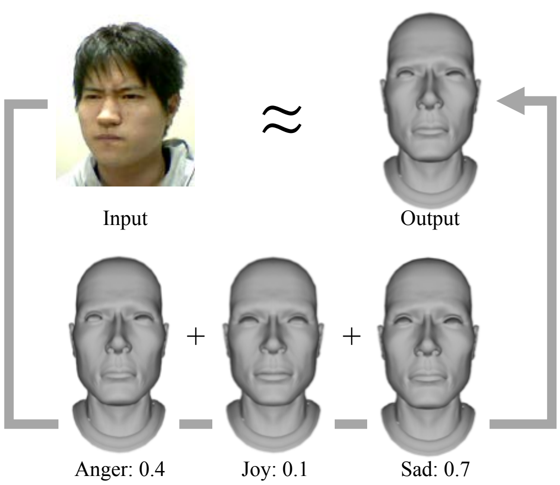

Figure 2. A corresponding 3DMM for a subject.

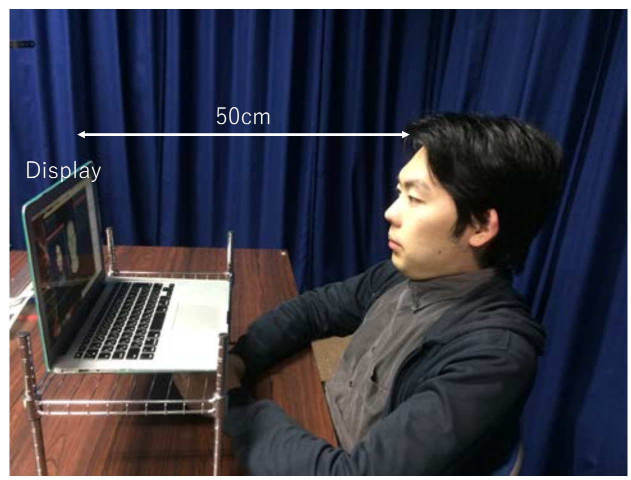

Figure 3. The experiment setup in this study.

proposed system uses blended emotive expressions of the models to find the most similar shape that matches a given facial image. This paper is organized as follows. Section II discusses known approaches to facial expression imitation. Section III introduces our proposed facial mimicry training system. Section IV shows the experiment results of the proposed system. Finally, Section V gives a short conclusion and highlights the most important outcomes of this paper.

II. RELATED WORK

3DMM is a well-established technique in computer graphics that produces expressive and plausible animations. This technique has been used to clone expressions from one face mesh to another. The cloning processes take two steps: determining surface points in the target correspond to vertices in the source model and transfer motion vectors from vertices of the source model to the target model. Sumner and Popović [10] proposed deformation transfer for triangle meshes, where the cloning process does not require the source and the target model to share a number of vertices or triangles. Figure 1 demonstrates deformations of faces [10] based on their expressions where the expression intensity varies from 0 to 1.0.

To fit 3DMM into a facial image, some points of 3DMM are associated to the corresponding landmark points in the facial image [11]. Extracting landmark points from facial images can be done using automated facial landmarks tools, such as Dlib library [12]. Those 2-dimensional (2D) landmarks are mapped into 3-dimensional (3D) using a 2D-to-3D registration method by referring to 3D facial points. The

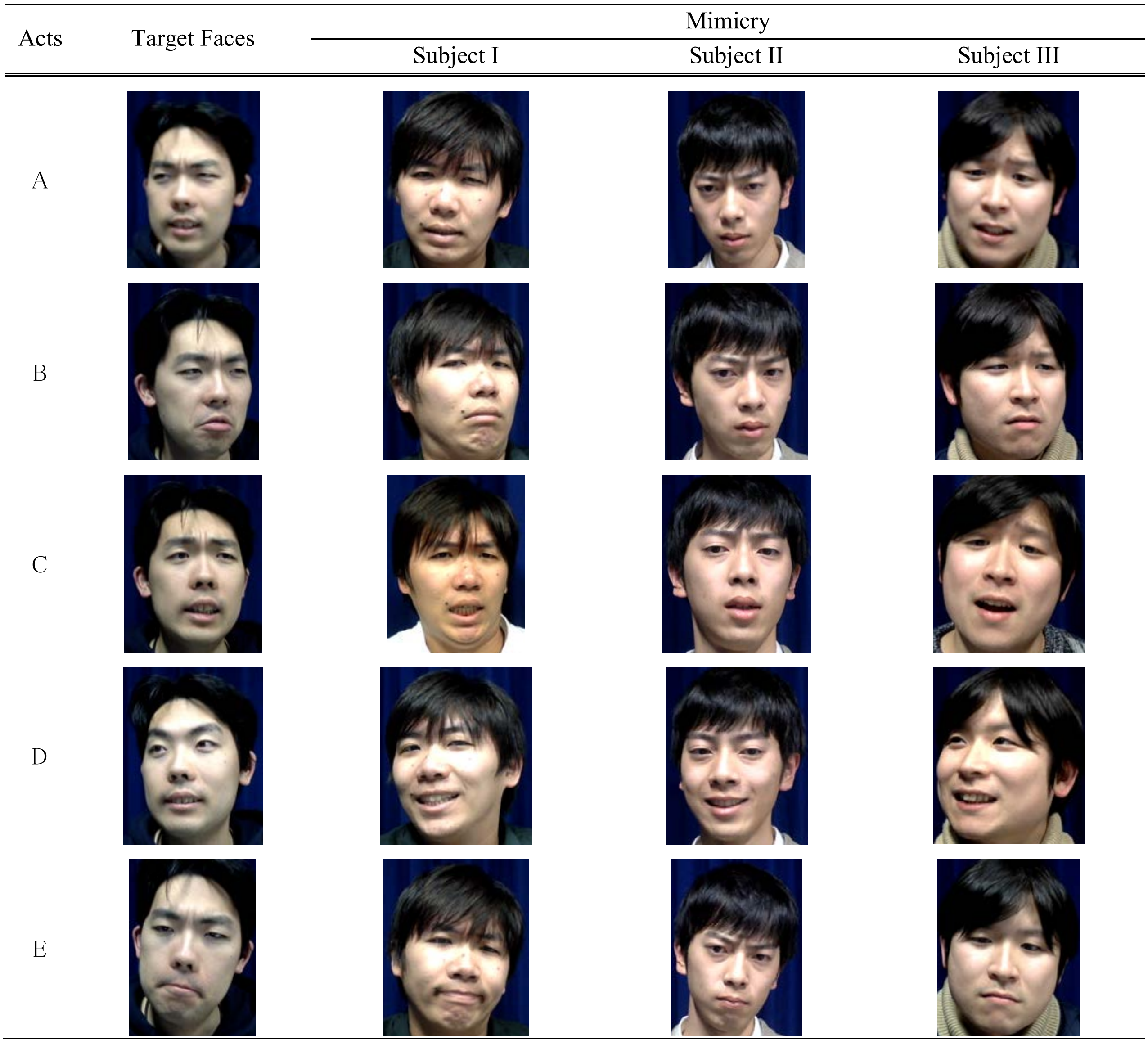

Figure 4. Results of mimicry training performed by three subjects.

resulted 3D landmarks are used for head-pose normalization. Figure 2 shows 3DMM imitating a subject's expression by blended emotive expressions.

III. Facial Mimicry Training

Our facial mimicry training system uses Dlib library to automatically annotate 68 landmarks from the facial image. These landmarks are associated to 3D face points [13] and then the head pose is calculated. SolvePnP library is used to solve the Perspective-n-Point (PnP) problem to perform 2D-to-3D registration [14].

Figure 3 shows the experimental setup in this study. The system uses a built-in webcam to capture the subject's face. The subject selects a target face to mimic from the database. While mimicking the target faces, subjects are instructed to adjust their head posture to match the target faces. When the subjects feel that they have precisely mimicked the target face, they press the "analyze" button to analysis the score for the mimicry. During the experiment, we did not specify the time required by each subject to mimic a target face. The resulted 3DMM for the subject is outputted along with the 3DMM for each emotive expression that makes up the result. The score for the mimicry is calculated as the correlation coefficient between the 3D points of the generated 3DMM for the subject and the target.

IV. Result

Three male subjects (mean age 21.7 years) were recruited for the experiments. All subjects agreed to participate and signed the consent forms, to allow their data to be used in publications of this research. Figure 4 shows the results of mimicry training performed by the subjects.

 ISBN: 978-1-61208-761-0

Table I to III show the correlation coefficients of the resulted mimics by the three subjects against the target faces. There are high correlations among the 3DMM of target faces and subjects' faces mimicking those target faces (values shown on gray background). Here, the correlation coefficients are above 0.98 for 3DMM of the target faces and their mimicries. When subjects were mimicking different faces, the correlation coefficients are below 0.94. These results show that, although most different places in the changes in facial expressions only occur at the upper part of the face (eyes and eyebrows) and the lower part of the face (lips), there is significant correlation between 3DMM and their mimicries.

V. Conclusion and Future Work

In this study, we have demonstrated that our self-learning-based facial mimicry training system is able to measure how close a person can mimic another person's facial expressions. By using this tool, users can train themselves to closely mimic someone's face interactively by referring to the expression intensity of each 3DMM constructing the blended 3DMM. In our further study, we will confirm the performance of the training system using a fine 3DMM that is generated from a large three-dimensional face dataset [15].

TABLE I. Correlation coefficients of the resulted mimics by the subject I against the target faces.

		Target Faces				
		A	B	C	D	E
Mimicries	A	0.994	0.908	0.894	0.913	0.873
	B	0.913	0.991	0.907	0.925	0.899
	C	0.881	0.891	0.987	0.926	0.923
	D	0.905	0.917	0.931	0.993	0.929
	E	0.87	0.891	0.923	0.924	0.997

TABLE II. Correlation coefficients of the resulted mimics by the subject II against the target faces.

		Target Faces				
		A	B	C	D	E
Mimicries	A	0.988	0.908	0.892	0.906	0.867
	B	0.913	0.994	0.906	0.923	0.896
	C	0.88	0.888	0.985	0.926	0.927
	D	0.894	0.91	0.924	0.989	0.925
	E	0.847	0.872	0.908	0.906	0.985

TABLE III. Correlation coefficients of the resulted mimics by the subject III against the target faces.

		Target Faces				
		A	B	C	D	E
Mimicries	A	0.997	0.912	0.899	0.915	0.877
	B	0.919	0.995	0.911	0.928	0.901
	C	0.895	0.903	0.998	0.935	0.93
	D	0.912	0.924	0.934	0.997	0.929
	E	0.866	0.888	0.921	0.921	0.996

References

[1] P. Ekman et al., "Universals and cultural differences in the judgments of facial expressions of emotion," *Journal of Personality and Social Psychology*, 53(4), pp.712-717, 1987.

[2] M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Measuring facial expressions by computer," Psychophysiology, Cambridge University Press, 36(2), pp. 253–263, 1999.

[3] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Measuring Facial Expressions by Computer Image Analysis," *Psychophysiology*, vol. 36, pp. 253-263, 1999.

[4] Y. L. Tian, T. Kanade, and J. F. Cohn, "Recognizing lower face action units for facial expression analysis." Proceedings - 4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000, 23(2), pp. 484–490, 2000.

[5] Y. Tian, T. Kanade, and J. F. Cohn, "Dual-State Parametric Eye Tracking," Proc. International Conference on Automatic Face and Gesture Recognition, pp. 110- 115, Mar. 2000.

[6] D. McDuff et al., "AFFDEX SDK: A cross-platform real-time multi-face expression recognition toolkit," Conference on Human Factors in Computing Systems, pp. 3723–3726, 2016. https://doi.org/10.1145/2851581.2890247

[7] M. Magdin, L. Benko, and Š. Koprda, "A case study of facial emotion classification using affdex," Sensors, 19(9), pp. 1–17, 2019. https://doi.org/10.3390/s19092140

[8] H. Miwa et al., "Effective emotional expressions with Emotion Expression Humanoid Robot WE-4RII - Integration of humanoid robot hand RCH-1," 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 3(4), pp. 2203–2208, 2004.

[9] K. Ito and Y. Nishimua, "An Interaction Analysis of System Usage: User-Instructor Interaction on the System named 'iFace'," *The Transactions of Human Interface Society*, 16(1), pp. 51–62, 2014. (in Japanese)

[10] R. W. Sumner and J. Popović, "Deformation Transfer for Triangle Meshes," *ACM Transactions on Graphics*, 23(3), pp. 399-405, 2004.

[11] H. Dai, N. Pears, W. Smith, and C. Duncan, "A 3D Morphable Model of Craniofacial Shape and Texture Variation," In 2017 IEEE International Conference on Computer Vision (ICCV), pp. 3085–3093, 2016.

[12] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014), pp. 1867–1874, 2014.

[13] X. Zhang. Face 3D Model File in CSV. [Online]. Available: https://xiaohaionline.com/wp-content/uploads/2017/10/average-face3d.csv [retrieved: March, 2020]

[14] X. X. Lu, "A Review of Solutions for Perspective-n-Point Problem in Camera Pose Estimation," In First International Conference on Advanced Algorithms and Control Engineering, pp. 1–8, 2018. https://doi.org/doi :10.1088/1742-6596/1087/5/052009

[15] B. Eggerr et al., "3D Morphable Face Models -- Past, Present and Future," ArXiv Preprint ArXiv:1909.01815, 2019.