Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Geometrical Feature-based Facial Expression Classification and Reproduction Method for Humanoid Robots

Xiaopan Kang, Manjia Su, Yihong Zhang,Haifei Zhu and Yisheng Guan
Biomimetic and Intelligent Robotics Lab(BIRL)
School of Electromechanical Engineering
Guangdong University of Technology
Guangzhou, Guangdong Province, China
ysguan@gdut.edu.cn

***Abstract*— Human robot interactions have become essential for many people in recent years. Facial expression reproduction is expected to be applied in human?robot interactions for conveying feelings. In this paper, geometrical features are defined to describe the generalized features of facial expressions, so that they can be used to recognize human expressions and replicate them in a robot. Using the geometrical features, a generalized distance between two expressions was proposed to classify expression types after all weights of geometrical features had been obtained. An additive actuation method based on the Blendshape model was also used to achieve the target expression with a certain accuracy. Finally, some simulation experiments were conducted to verify these methods using BU-4DFE, including the visualization and analysis of geometrical features, the calculation of weight and its power, and the clarification and actuation of a target expression. The visualization of geometrical features indicated the intuitive performance, while the analysis of the features confirmed that they follow certain geometrical rules. All weights and the power were calculated to confirm the categorization of expressions. The classification rates and relative errors proved the reliability of the geometrical features and the calculated weight results. Finally, the results of additive actuation of feature points provided a significant reference for designing the actuation points of a robot face, which depends on how high an accuracy of expressions we want to achieve. The methods and results of this paper can be applied to developing a practical robot face.**

***Index Terms*— geometrical features, facial expressions, classification, reproduction, humanoid robots.**

I. Introduction

With the ongoing development of manufacturing technology and computer science, more and more humanoid robots are entering into peoples everyday lives, not only looking like humans but also acting like humans. They are in line with human aesthetics and will be able to accurately convey information and emotions. Concerning face-to-face communication, Mehrabian [1] indicated that only 7% of the message is from linguistic language; 38% from intonation and 55% is transmitted by facial expressions. This implies that facial expression is a major modality for humanoid robots. Thus, a humanoid robot that can produce vivid facial expressions is highly desirable. The facial expressions of same humanoid robots have been studied since 1992 by Tomohiko Umetsu at Waseda University [2]. The series of humanoid robots developed by Waseda University can make nine expressions and they have other functions, like smell, touch, and response to light. SAYA [3], made by Tokyo University of Science, could recognize six facial expressions and reproduce them via a pneumatic muscle actor. Beautiful Announcer [4] was used in 2018, with lifelike facial expressions and communication with people. F&H Robot [5], in development since 2004, has facial expressions and interactive functions, like speech recognition and dialogue.

The reproduction of facial expressions in humanoid robots is based on the facial action coding system (FACS), which was established by Ekman [6]. He divided the human face into 46 action units according to the formation mechanism of facial expressions. Furthermore, he observed and explored the corresponding relationship between action units and six basic facial expression (happy, sad, disgust, angry, surprise, and fear). When a humanoid robot needs to make a facial expression, the controller drives the relevant action units. Then the expression occurs in the robots face. Indeed, it is an effective way to reproduce facial expressions in humanoid robots, but there is no numerical criterion to measure whether the facial expression is similar to a humans face. Naturally, a humanoid robot cannot adjust the volume of the controller to reproduce a facial expression exactly due to the lack of precise numerical results when comparing a humanoid robots facial expression with a humans.

To make a humanoid robot reproduce a facial expression convincingly, it is necessary to provide a quantitative description of the facial expression, classification method, and actuation method. Not only a qualitative description, but also the driving points are calculated in order to decrease the numerical difference between the current and target expressions. This is a major motivation for us with this paper.

*This research was supported by the National Natural Science Foundation of China (Grant Nos. 51975126, 51705086, 51605096), the Frontier and Key Technology Innovation Funds of Guangdong Province (Grant Nos. 2017B050506008, 2017B09091000,2 and 2019B090915001), and the Program for Guangdong Yangfan Introducing Innovative and Entrepreneurial Teams (Grant No. 2017YT05G026).

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

II. Tools and Framework

A. Database of Facial Expressions

To make a humanoid robot reproduce real human facial expressions, the relationship between them should be established. Previously, it was necessary to obtain data on real human facial expressions. The database BU-4DFE [7] is used in this paper, which contains 101 subjects. For each subject, there are six model sequences showing six prototypical facial expressions (anger, disgust, happiness, fear, sadness, and surprise). We can get the original data of any subject and any facial expression by the following toolkit.

B. Database of Facial Expressions

The quantitative?description of facial expressions and numerical difference between a humanoid robots facial expression and a humans are to be proposed. The tool we used to get the initial data on facial expressions is OpenFace, based on the CVPR 2015 paper [8].

This toolkit can perform a number of facial analysis tasks. In this paper, we use the original data file, which contains the three-dimensional coordinates of 68 feature points defined in OpenFace, concentrating on the eyes, mouth, eyebrows, and facial contours.

C. Framework

The method proposed in this paper can be divided into three parts, including the geometrical features, the classification, and the actuation of the facial expression. The geometrical features can use vectors to define a facial expression. Then the different facial expressions are classified according to numerical difference based on features formerly defined in the database. Based on the known facial expressions, it is actuated to perform using the Blendshape Model [9]. The whole flowchart of such a framework is shown in Figure 1.

III. Geometrical Feature

OpenFace outputs 68 3D feature points, but they are not distinct enough to describe a facial expression. One reason is that a humanoid robots face and a humans face are not identical, even if they have the same basic facial expression. Thus, we propose some geometrical features of facial expression, including shape features and deformation features, to present more information in global and local detail.

A. Shape Feature

Different facial expressions that consist of different facial areas, such as the mouth, eyes, nose, and eyebrows, should be classified because they present significantly different features. Shape features are chosen to define the shape deformation, which are shown by the chain code [10] in Figure 2 and Table I. They have various geometrical features, like line drawings, planar curves, contours, and region shapes. The chain code

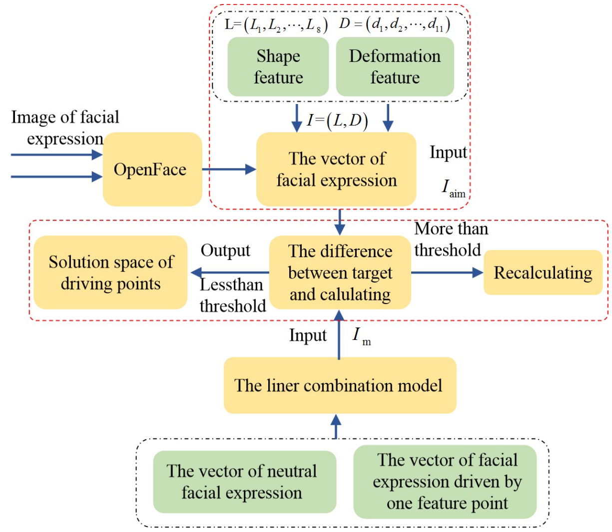


Fig. 1. The framework to define, classify, and actuate a facial expression.

is calculated using the angles between the line segments and a horizontal line.

In this paper, we choose 48 direction chain codes that include 48 feature points to describe the shape feature. They are defined in relation to the horizontal?and vertical motion of all 68 feature points from the six chosen facial expressions. The value of the chain code of every feature point k_i is

$$k_{\mathrm{i}} = \frac{180\theta}{\pi} \tag{1}$$

$$\theta = \begin{cases} \frac{\pi}{2}, & X_{i+1} = X_i, Y_{i+1} \geq Y_i \\ \frac{3\pi}{2}, & X_{i+1} = X_i, Y_{i+1} < Y \\ \arctan\frac{Y_{i+1}-Y_i}{X_{i+1}-X_i}, & X_{i+1} > X_i, Y_{i+1} \geq Y_i \\ \arctan\frac{Y_{i+1}-Y_i}{X_{i+1}-X_i} + 2\pi, & X_{i+1} > X_i, Y_{i+1} < Y \\ \arctan\frac{Y_{i+1}-Y_i}{X_{i+1}-X_i} + \pi, & X_{i+1} < X_i \end{cases} \tag{2}$$

where (X_i, Y_i) is the ith feature points and θ is the angle between the ith line segment $p_i p_{i+1}$ and the horizontal line. We can calculate the value of chain code $(k_1, k_2, \cdots, k_m)$ of the facial area with m feature points. Their first difference l_i is

$$l_i = \begin{cases} k_i - k_{i+1}, & i = 1, 2, \cdots, m-1 \\ k_i - k_0, & i = m \end{cases} \tag{3}$$

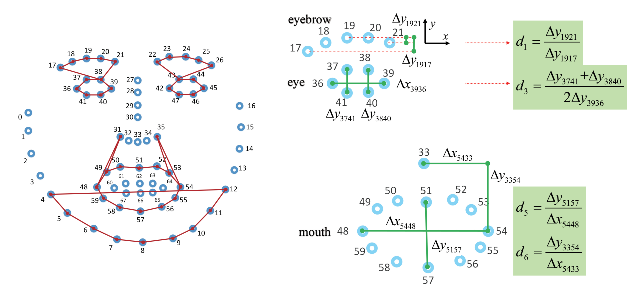
Non-negative operation is applied to l_i and it becomes

$$l_i = \begin{cases} l_i + 360 - 1, & i < 0 \\ l_i, & i \geq 0 \end{cases} \tag{4}$$

The vector of the chain code used to describe the shape feature is $L = l_1, l_2, \cdots, l_m$. In this paper, we choose the shape features of the mouth, eye, lower jaw, and eyebrow. They are divided into eight groups, as shown in Table I, and the shape feature is shown in Figure 2(a).

TABLE I
FEATURE POINT GROUPS OF DIFFERENT FACIAL AREAS

Facial Area	Sequence of Feature Points	Number of Feature Points
Mouth l_1	48-49-50-51-52-53-54 -55-56-57-58-59-48	12
Left eye l_2	36-37-38-39-40-41-36	6
Right eye l_3	42-43-44-45-46-47-42	6
Left eyebrow l_4	17-18-19-20-21-38-17	6
Right eyebrow l_5	22-23-24-25-26-43-22	6
Jaw l_6	4-5-6-7-8-9-10-11-12-4	9
Left side of mouth l_7	48-49-31-48	3
Right side of mouth l_8	54-53-35-54	3



(a) The shape feature of eight groups including areas of the mouth, eye, eyebrow, and jaw. (b) The deformation featuresła total of 10 based on the definition.

Fig. 2. The actuation performance with single and multiple feature points.

B. Deformation Features

Although shape features are used to describe the relationship of neighboring feature points, they cannot reflect the deformation of facial area, such as contraction and expansion. Thus, we define the deformation features (Figure 2(b)) of facial areas, such as the extent of the mouth opening and closing. These are based on the distribution of facial expression muscles.

We chose 10 deformation features to present a unique combination of muscles used in different facial expressions. They are defined as follows:

- Left eyebrow (the value increases if the left eyebrow lifts):

$$d_1 = \frac{Y_{19} - Y_{21}}{Y_{19} - Y_{17}} \tag{5}$$

- Right eyebrow (the value increases if the right eyebrow lifts):

$$d_2 = \frac{Y_{24} - Y_{22}}{Y_{24} - Y_{26}} \tag{6}$$

- Left eye (the value increases if the left eye opens wide):

$$d_3 = \frac{(Y_{37} + Y_{38}) - (Y_{40} + Y_{41})}{2\,(X_{39} - X_{36})} \tag{7}$$

- Right eye (the value increases if the right eye opens wide):

$$d_4 = \frac{(Y_{43} + Y_{44}) - (Y_{46} + Y_{47})}{2\,(X_{45} - X_{42})} \tag{8}$$

- Left corner of mouth (the value increases if the left corner of the mouth rises):

$$d_5 = \frac{Y_{57} - Y_{51}}{X_{54} - X_{48}} \tag{9}$$

- Left corner of mouth (the value increases if the left corner of the mouth rises):

$$d_6 = \left|\frac{X_{33} - X_{48}}{Y_{33} - Y_{48}}\right| \tag{10}$$

- Right corner of mouth (the value increases if the right corner of the mouth rises):

$$d_7 = \left|\frac{X_{33} - X_{54}}{Y_{33} - Y_{54}}\right| \tag{11}$$

- Left cheek (the value increases if the left cheek lifts):

$$d_8 = \left|\frac{X_{31} - X_{48}}{Y_{31} - Y_{48}}\right| \tag{12}$$

- Right cheek (the value increases if the right cheek lifts):

$$d_9 = \left|\frac{X_{35} - X_{54}}{Y_{35} - Y_{54}}\right| \tag{13}$$

- Jaw (the value increases if the jaw drops):

$$d_{10} = \frac{2Y_8}{Y_4 + Y_{12}} \tag{14}$$

In order to make the deformation features have scale invariance and comparability, the 10 deformation features should have a standard distance between them. Since the distance between the inner corner of the two eyes is the same in different kinds of facial expressions, all deformation features become

$$d_i = \frac{d_i}{\sqrt{(Y_{42} - Y_{39})^2 + (X_{42} - X_{39})^2}} \tag{15}$$

The deformation features convey enough information about expression intensity-related facial expression muscles. So, the facial expression can be described using both shape features and deformation features, as in

$$I = (L, D) \tag{16}$$

where shape features are $L = (l_1, l_2, \cdots, l_8)$ and deformation features are $D = (d_1, d_2, \cdots, d_{10})$.

IV. Classification and Actuation

A. Classification for Different Facial Expressions

Since a vector $I=(L,D)$has been used to represent the facial expression, we can calculate the generalized distance between the facial expressions of different humans. Similarly, the generalized distance can be applied for mapping the difference between a human expression and a robots. Supposing that we are going to calculate the generalized distance between I and F, where they are

$$\begin{array}{c} I=\left(L_1^I, L_2^I, \cdots, L_s^I, D_1^I, D_2^I, \cdots, D_{10}^I\right) \\ F=\left(I_1^F, L_2^F, \cdots, I_s^F, D_1^F, D_2^F, \cdots, D_{10}^F\right) \end{array} \tag{17}$$

the difference of each feature $d_i(I,F)$ is as follows:

$$d_i(I,F)=\begin{cases} 2\sum_{k=1}^{m_i}\frac{\left|L_{ik}^l-L_{ik}^F\right|}{L_{ik}^l+L_{ik}^F}, & 1\leq \mathrm{i}\leq 8 \\ 2\frac{\left|D_{i-p}^l-D_{i-p}^F\right|}{D_{i-p}^l+D_{i-p}^F}, & 8\leq i\leq 18 \end{cases} \tag{18}$$

Then the generalized distance of the two expressions is

$$S(I,F)=\sum_{i=1}^{18} d_i^2(I,F) \tag{19}$$

Equation (19) shows that every feature is equally important. In fact, different features make different contributions to performing a facial expression. Generally, the contribution of the eyebrows, eyes, and mouth is the most significant, while the cheek and jaw are less significant, and the nose can even be ignored in most facial expressions. The deformation amplitude of the mouth area is larger than that of the eye area. The deformation amplitude of the eyebrow area is smaller than that of the mouth and cheek area. In order to distinguish the contributions of different features, we modify Equation (19) as follows:

$$S(I,F)=\sum_{i=1}^{18} \omega_i^\sigma d_i^2(I,F) \tag{20}$$

where $w_i\geq 0,(i=1,2,\cdots,p+q)$ and $\sum_{i=1}^{18} w_i=1$, representing the weight of the corresponding feature. $\sigma:\sigma\geq 0$ is the power of the current weight.

To obtain the weight and the weight effect coefficients σ, we chose the same kind of facial expressions from BU-4DFE. After calculating the values of the shape features and deformation features, we used the generalized distance as a goal function to evaluate the two expressions. If they are close enough, they are classified into the same kind. If not, they are different expressions.

$$\begin{aligned} J_{w,\sigma} &= \sum_{N=1}^{C_{101}^2}\omega_i^\sigma d_i^2(I,F)=\sum_{N=1}^{C_{101}^2}\sum_{i=1}^{18}\omega_i^\sigma d_i^2\left(I_N,F_N\right) \\ &= \sum_{i=1}^{18}\omega_i^\sigma\sum_{N=1}^{C_{101}} d_i^2\left(I_N,F_N\right) \\ \text{s.t.} & \sum_{i=1}^{18} w_i=1 \end{aligned} \tag{21}$$

To calculate the weight and the weight effect coefficients, it is necessary to solve (21). In this paper, we attempt different algorithms, linear programming $\sigma=1$, nonlinear programming $\sigma=1$, and quadratic programing $\sigma=2$, to find the best solution.

B. Additive Actuation of Feature Points

Our goal was to make a robot reproduce the facial expressions of a human, so that the generalized distance can be regarded as a target. Then we proposed an additive actuation method to make a robot face reproduce the target expression. The facial expression of a robot face can be formed by combining the feature points from a neutral to a target facial expression, based on the Blendshape Model [9]. The linear additive combination of blend shapes is

$$I_{aim}=I_{neural}+\sum_{m=1}^{68}\alpha_m I_m \tag{22}$$

where I_{aim} is the vector of target expression, I_{neutral} is the vector of neutral expression, and I_{m} is the vector of expression driven by one feature point. α_{m} is the Boolean factor of a feature point. If $\alpha_{\mathrm{m}}=1$, $\alpha_{\mathrm{m}}=0$the feature point is enabled; if , the feature point is disabled.

The vector I_{m}, a description of a facial expression driven by one feature point, can be obtained by scattered data interpolation [11]. When given a set of known displacements away from the original position at every constrained point, we construct a function that gives the displacement for every unconstrained point. In this way, we can get the vector driven by one feature point. Then we rank them based on the generalized distance between two expressions. Afterwards, a certain dimension vector combined with such feature points actuates the current expression to the target expression. The dimension number of this vector is determined by setting a generalized distance threshold. This vector, including the feature points, would be applied to design a practical robot face.

V. SIMILATION and Verification

We conducted experiments using BU-4DFE, including the visualization and analysis of geometrical features, the calculation of weight and its power, the clarification, and the actuation for a target expression.

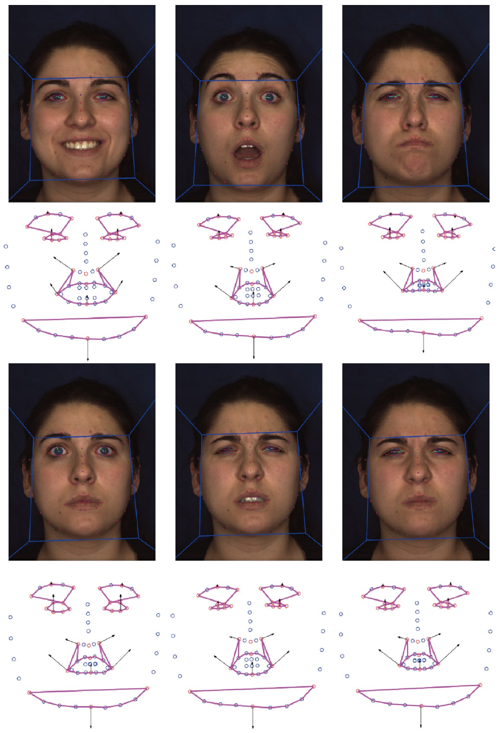


Fig. 3. The visualization of geometrical features, including Happy, surprised, sad, Fear, disgusted, and angry expressions.

A. Visualization and analysis of geometrical features

According to the defined geometrical features, they were visualized and drawn in Figure 4. In Figure 4, all shape features were pictured using red lines, based on the 68 feature points after recognition by OpenFace. while all deformation features were presented by drawing arrows in their feature directions. The magnitude of the arrow showed the intensity of the deformation feature. We showed all the facial expressions of one subject in BU-4DFE.

The visual figures showed the distortion of shape features and the intensity of deformation features. The results verified the accuracy of the geometrical features.

To analyze the reproduction of a neutral facial expression from the others, the reliability of the geometrical features was verified by comparing the neutral facial expressions and the others. In Figure 5, we give an example of the deformation feature curves from a neutral to a happy expression. We discover some rules:

- The deformation features remain roughly equal during the process of 101 subjects performing a happy expression. Thus, we used the average value to represent the value of such a deformation feature. Notably, we print-screened about 400 figures from 101 subjects happy videos, about four figures per subject, to enlarge the base number.
- Figure 5b, the difference of some deformation features (e.g., D2) between the neutral and happy expressions is small, while the difference of some other deformation feature (e.g., D4) is large in Figure 5c. The deformation features with strong differences between two expressions are regarded as vital features. This phenomenon indicated that expressions must be clarified using these geometrical features.

TABLE II
CLARIFICATION RATE AND RELATIVE ERROR OF EXPRESSIONS

Expression	Ha.	Sa.	An.	Fe.	Di.	Su.
Recognition rate	90.9	86.21	95.24	90.62	95.10	92.51
Relative error	9.10	13.79	4.76	9.38	4.90	7.49

B. Weight and clarification

In order to obtain the weights of all kinds of expressions, we addressed an optimized problem according to Equation (20). First, we used one kind of expression from 101 subjects to obtain the number () of generalized distances . Then the weight and its power became the variables of the function . After the optimized problem was solved, a set of weights and the power could be obtained. Similarly, all weights for different expressions could also be obtained. Using the weights, we can clarify whether a given expression belongs to a certain category. To present the clarification rate and relative error, 707 expressions (seven kinds of expressions of one subject 101 subjects) were compared with the expression patterns in Table II. Finally, we obtained the results in Table II., which proves the strong reliability of the geometrical features and the calculated results of weight.

C. Actuation

The additive actuation based on the Blendshape model in (21) makes a robot face system with feature points able to reproduce a target expression. In Figure 6a, we give an instance where a feature point drives the whole face to a target position, with the other points deforming compatibly according to the Blendshape model. The motion driven by this feature point has a contribution (the generalized distance) that makes the robot face tend to the target expression. Then all feature points were ranked according their strength of contributions. They would be added to the actuation point set one by one following the contribution sequence. Once we determined to make the robot face achieve 39% of the target expression, we got the actuation point set in Figure 6b (left image). If we want a better performance, to achieve 95% of the target motion, the actuation point set became that in Figure 6b (right image). Notably, the actuation points were not symmetrical about the vertical centerline. This result provided a significant reference for designing the actuation

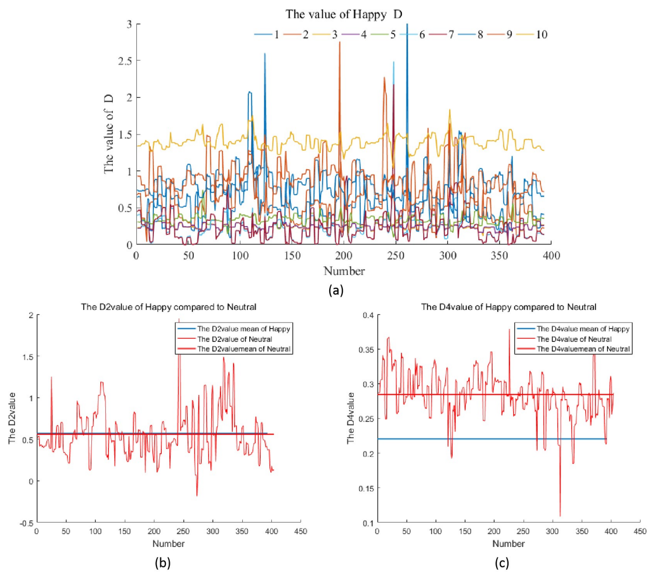
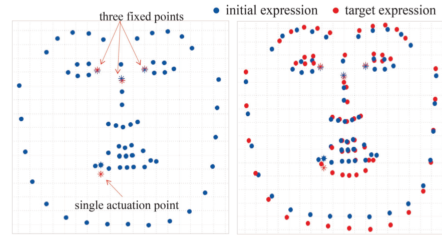
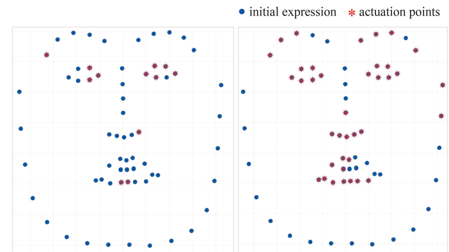


Fig. 4. The deformation features. (a) The change trend of the 10 deformation features is approximately the same in the same facial expression. (b) The change trend of the eyebrow is small, so it is not an important deformation feature in the shift from neutral to happy. (c) The change trend of the eye is obvious as it becomes smaller than in a neutral expression, so it is important in the shift from neutral to happy.



(a) The actuation by a single point.



(b) The actuation point sets that achieve 39% and 95% accuracy

Fig. 5. The actuation performance with single and multiple feature points.

points of a robot face, depending on how high an accuracy of expressions we want to achieve.

VI. CONCLUSIONS

This paper aimed to reproduce facial expressions in humanoid robots. Previously, we defined geometrical features, including shape and deformation features, to describe the generalized features of a facial expression. Using geometrical features, we were able to categorize the expressions into types based on the weights of all features and the convergence of the generalized distance between two expressions. Finally, an additive actuation method based on the Blendshape model was applied to achieve the target expressions with a given accuracy. These were the main methods for reproducing facial expressions on a robot face.

Some experiments were conducted to verify these methods using BU-4DFE, including the visualization and analysis of geometrical features, the calculation of weight and its power, and the clarification and actuation of a target expression. The visualization of the geometrical features indicated the intuitive performance of the represented feature, and the analysis of the geometrical features showed that they followed a common set of rules. All weights and the power were calculated, so that they could be used to clarify kinds of expressions. After the statistical calculation, the clarification rates and relative errors proved the strong reliability of the geometrical features and calculated results of weight. Finally, the results of additive actuation of feature points provided a significant reference for designing actuation points on a robot face, which depends on how high an accuracy of expressions we want to achieve.

The methods and results of this paper have been proposed and verified theoretically. However, the next challenge is to develop a practical robot face system to verify and achieve the reproduction method. This work is ongoing for our research team.

REFERENCES

[1] Mehrabian. Communication without words. Journal of Psychology today. 1968, vol.2, no.4, pp.53-56.
[2] Miwa, Hiroyasu , H. Takanobu , and A. Takanishi . Human-like Head Robot WE-3RV for Emotional Human-Robot Interaction. Romansy 14. Springer Vienna, 2002.
[3] Hashimoto, Takuya, et al. Development of the Face Robot SAYA for Rich Facial Expressions. International Joint Conference SICE-ICASE, IEEE 2006.
[4] https://www.youtube.com/watch
[5] Weiguo, W. , Qingmei M., and Yu W. Development of the humanoid head portrait robot system with flexible face and expression. International Conference on Robotics & Biomimetics, IEEE 2005.
[6] Ekman, Paul, and Rosenberg E. L. What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS), 2nd ed. 2005.
[7] Lijun Y., et al. A high-resolution 3D dynamic facial expression database. 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008.
[8] Schroff F., Kalenichenko D., Philbin J. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - FaceNet: A unified embedding for face recognition and clustering. 2015: 815-823.
[9] Wilbers, Freerk, and C. Ishi, et al. A blendshape model for mapping facial motions to an android. IEEE/RSJ International Conference on Intelligent Robots & Systems. 2007.
[10] Trucco, Emanuele. Computer and Robot Vision. 1995.
[11] Frdric P., et al. Synthesizing realistic facial expressions from photographs. Siggraph'98: Proc. Conference on Computer Graphics and Interactive Techniques 1998.