Automatic Facial Expression Learning Method Based on Humanoid Robot XIN-REN

Fuji Ren, *Senior Member, IEEE*, and Zhong Huang

***Abstract*—The ability of a humanoid robot to display human-like facial expressions is crucial to the natural human–computer interaction. To fulfill this requirement for an imitative humanoid robot, XIN-REN, an automatic facial expression learning method is proposed. In this method, first, a forward kinematics model, which is designed to reflect nonlinear mapping relationships between servo displacement vectors and corresponding expression shape vectors, is converted into a linear relationships between the mechanical energy of servo displacements and the potential energy of feature points, based on the energy conservation principle. Second, an improved inverse kinematics model is established under the constraints of instantaneous similarity and movement smoothness. Finally, online expression learning is employed to determine the optimal servo displacements for transferring the facial expressions of a human performer to the robot. To illustrate the performance of the proposed method, we conduct evaluation experiments on the forward kinematics model and the inverse kinematics model, based on the data collected from the robot's random states as well as fixed procedures by animators. Further, we evaluate the facial imitation ability with different values of the weighting factor, according to three sequential indicators (space-similarity, time-similarity, and movement smoothness). Experimental results indicate that the deviations in mean shape and position do not exceed 6 pixels and 3 pixels, respectively, and the average servo displacement deviation does not exceed 0.8%. Compared with other related studies, the proposed method maintains better space–time similarity with the performer, besides ensuring smoother trajectory for multiframe sequential imitation.**

***Index Terms*—Expression mapping, forward kinematics model, humanoid robot, inverse kinematics model, movement similarity, movement smoothness.**

I. Introduction

HUMANOID robots, which are designed by using artificial muscles, deformable skin, and motion control systems, are widely used to assist the humans in their daily chores and day-to-day activities, such as newscasting, teaching, and guidance [1], [2]. Facial expressions, which are well recognized to be critical to conveying motivational states and emotional feelings of humans, are essential for humanoid robots. Hence, generating robot expressions similar to those of humans, under the servo constraints and hardware limitations, is essential to promote natural, trustworthy, human–robot interactions [2]–[4].

With the development of expression mapping technologies in the facial animation domain [5], [6], the performance-driven or learning-from-observation techniques, which map human expressions to 2-D or 3-D virtual avatars, provide a reference for transferring facial expressions from humans to humanoids [7], [8]. However, the process of generating facial expression for mechanically operated humanoid robots is more challenging than generating for virtual characters. And, three reasons can be cited for this. First, humanoid robots cannot be embedded with enough servos because of head space constraints and servo technologies [9], [10]. Humanoid robots have fewer movement joints and degrees of freedom (DOFs) than virtual avatars and humans, which have 268 voluntary muscles to generate facial expressions. Second, servo system designed for humanoid robots do not follow the facial action coding system, which defines 44 facial action units (AUs) for a facial animation of virtual avatars. Thus, generating facial expressions for humanoids by finding correspondences between servos and AUs is difficult [11]. Third, the multiservo control systems lag far behind the current graphical rendering techniques. To overcome these constraints, some methods, such as key-value store [12], polynomial fitting [13], and neural network [14], have been proposed to determine the optimal servo displacements needed to maximize expression similarity, based on inverse kinematics model. However, there has not been enough research on the smoothness of expression imitation, which is necessary for humanoids to reproduce natural and less hardwired expressions.

To make up for this deficiency, we propose an automatic expression learning method to map facial expressions of a human performer to a humanoid robot. First, 48 feature points, which move with the servo displacements, are tracked by an active appearance model (AAM) algorithm [15]. The forward kinematics model, which is built to reflect nonlinear mapping relationships between servo displacements and feature point positions, is converted into the linear relationships between the mechanical energy of servo displacements and the potential energy of feature points, based on the energy conservation principle. Second, an inverse kinematics model is designed under the constraints of instantaneous similarity and movement smoothness, based on the forward kinematics and the trajectory prediction models. In addition, the exterior point penalty function method [16] is employed to solve the multiconstraint nonlinear optimization. Finally, an automatic expression learning process is established to determine the optimal servo displacements for transferring

Manuscript received October 13, 2015; revised February 11, 2016, April 5, 2016, and June 17, 2016; accepted July 27, 2016. Date of publication August 30, 2016; date of current version November 11, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61432004 and Grant 61472117, in part by the Beijing Advanced Innovation Center for Imaging Technology (No. BAICIT-2016012), and JSPS KAKENHI Grant (No. 15H01712), and in part by the Natural Science Research Project of the Education Department of Anhui Province (No. AQKJ2015B013). This paper was recommended by Associate Editor J. Han. *(Corresponding author: Zhong Huang.)*

F. Ren is with the Faculty of Engineering, University of Tokushima, Tokushima 770-8501, Japan, and also with the School of Computer and Information, Hefei University of Technology, Hefei 230009, China (e-mail: ren@is.tokushima-u.ac.jp).

Z. Huang is with the School of Computer and Information, Hefei University of Technology, Hefei 230009, China, and also with the School of Physics and Electronic Engineering, Anqing Normal University, Anqing 246000, China (e-mail: huangzhong3315@163.com).

Digital Object Identifier 10.1109/THMS.2016.2599495

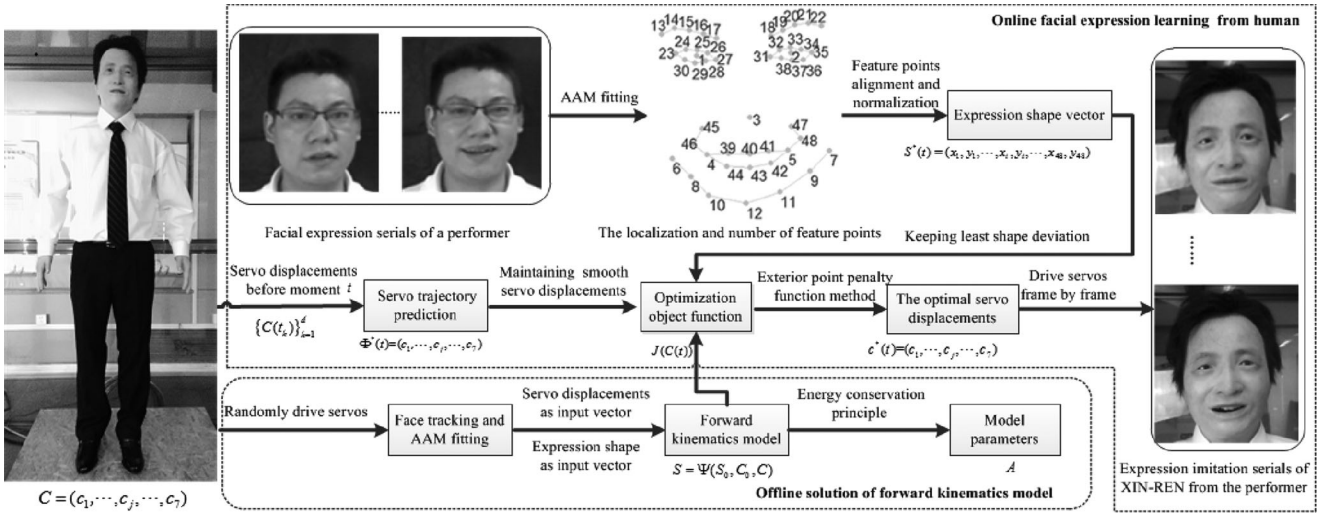


Fig. 1. Outline of the automatic facial expression learning method from a human performer.

facial expressions from the human performer to the robot, as well as maintaining a smooth trajectory under the inverse kinematics model. Fig. 1 shows the outline of the proposed method.

The following are the main contributions of this study.

1) The forward kinematics model, which maps the servo displacement space into expression shape space, is converted into the linear relationships between the mechanical energy of the servo displacement vectors and the potential energy of the expression shape vectors, based on the energy conservation principle. This strategy not only decouples the nonlinear relationships between servo displacements and feature point positions, but also provides requisite prior knowledge for establishing an inverse kinematics model. In addition, the feature points, tracked by the AAM algorithm, are noncontact and nonwearing [17]. Therefore, this nonwearable way is more conducive to collect training data for solving the forward kinematics model and improving human–computer interaction.
2) Different from the existing inverse kinematics solutions, which focus only on movement similarity, an improved inverse kinematics model is proposed under the constraints of instantaneous similarity and movement smoothness. This ensures shape consistency with the target shape vector, and maintains smooth movement trajectories for curbing the zigzag effect, besides protecting servo hardware and facial silicone skin. In addition, an adjustable factor is introduced into the inverse kinematics model to weight instantaneous similarity and movement smoothness.
3) Compared with the time-consuming manual setting of servo displacements, the proposed online expression learning algorithm can better reflect the inherent characteristics of human facial dynamics and reproduce natural and less hardwired robot expressions.
4) The rest of this paper is organized as follows: Section II summarizes the related work; Section III describes the humanoid robot XIN-REN; Section IV analyzes the forward kinematics model and the coefficient matrix solution; Section V describes the inverse kinematics solution; Section VI provides offline forward kinematics solution and online expression learning method; Section VII evaluates the forward kinematics model and the automatic facial expression learning method and Section VIII concludes the paper.

II. Related Work

Humanoid robots are required not only to have human-like appearance, but also to possess the ability of emotional interaction [18], [19]. Some research results show that robots with physical presence and facial expressions, similar to those of humans, can enable humans remember more interaction details, and provide more engaging and credible information than virtual characters generated by graphical rendering techniques [2]. Therefore, studies on humanoid robots with facial expressions are indispensable for effective interaction between humans and robots [2], [3]. But, several of the current humanoid robots, which have human-like appearance, such as Philip K. Dick [20], EveR series [3], Geminoid F [14], can only perform predefined and limited facial expressions, because they cannot replicate all human facial muscles and skeletal movements of the head [18]. These unbalances between appearance and movement give an eerie impression [4], [21]. A common method to generate facial expressions for humanoid robots is to perform several predefined target servo displacements based on a key-frame technology [17], which is used for cartoon animations. This method is simple to implement, but it can build only exaggerated expressions. The fidelity of facial expressions depends on the skill level of the animators; besides, the manual procedure is time consuming [7]. To generate expressions similar to those of the humans, and to maintain proper motions and timing, the straightforward method for humanoid robots is through imitation from humans [22], [23].

Research on a robot expression imitation is divided mainly into the expression pattern imitation and expression detail imitation. Expression pattern imitation is employed to replicate the same expression categories as those of a performer, under the constraint of fewer servos [24], [25]. By subjectively

observing the relationships between servos and AUs, Shayganfar *et al*. [11] propose a general strategy to generate expression patterns, such as happy, surprise, and sad. As the corresponding relationships between AUs and servos are not evident, the produced expression patterns are limited; besides, their fidelity needs further improvement. Ahn *et al*. [26], [27] propose a facial muscle control method to generate all expression categories by imitating the facial muscle mechanism of humans, and discuss how to prevent damage to servos and outer skin. Using the same facial muscle mechanism, Tadesse and Priya [9], [28] dissect geometric relationships between mimetic muscles and mechanical characteristics of servos, and realize facial transformation across different expression patterns by a graphical analysis.

Unlike the expression pattern imitation, which is used mainly for mechanical-looking robots, such as NAO [29], HanSaRam [30], Kismet [31], and KOBIAN-R [13], the expression detail imitation focuses on generating a similar expression intensity and movement trajectory. It not only sets higher standards for servo structures, DOFs, and response delay, but also demands more stringent requirements for real time. Wilbers *et al*. [17] propose an expression-learning method by transferring facial motions of a performer to the android Repliee Q2 for TV newscasting [32]. Employing blend-shape models, the Repliee Q2 can demonstrate realistic facial motions with a human performer. Jaeckel *et al*. [7] introduce an expression-mapping technology into the expression detail imitation system for humanoids. They establish a mapping model between facial deformation of a performer and 34 servos of the robot Jules, and use partial least squares to achieve expression imitation. The methods of Wilbers *et al*. [17] and Jaeckel *et al*. [7] are established based on the linear relationships between facial servo displacements and facial feature points. However, the linear models work well only when the servos and feature points coincide [14]. As the linear model does not sufficiently reflect the complexity of servo control system, researchers have introduced some nonlinear models into the inverse kinematics solvers. Magtanong *et al*. [33] use a back-propagation (BP) neural network to build forward kinematics model and solve inverse kinematics by optimization. Employing a feed-forward neural network and genetic algorithm, Habib *et al*. [20] put forward a learning method to transfer some facial features of humans to the android P. K. Dick. Trovato *et al*. [13] use three power polynomials to fit the nonlinear relationships between the servo displacements and the robot facial cues designed by an animator. By selecting appropriate combinations of facial cues, the method can produce numerous natural expressions for the humanoid robot KOBIAN-R. However, both pattern imitation and detail imitation focus more on the similarity of imitation more than on the smoothness of imitation.

III. Humanoid Robot Xin-Ren

To investigate human–computer interaction with emotions, an imitative humanoid robot, XIN-REN, has been developed by author Ren *et al*. With advanced servo control technology, XIN-REN can replicate major facial muscle actions to display various facial expressions. Fig. 2 shows its head structure, which is made

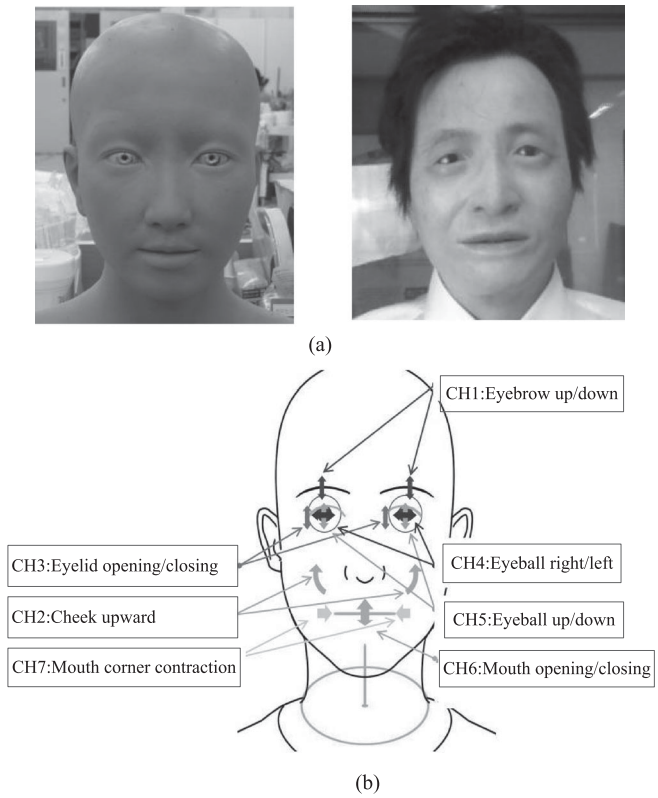


Fig. 2. Head structure of the robot XIN-REN. (a) XIN-REN's head structure and human-like appearance made of silicon jelly (b) Seven controllable head servos of XIN-REN

of silicon jelly to impart human-like appearance. Meanwhile, seven servos, driven by a pneumatic pump for safe and smooth interaction, are controlled by sending target servo displacements from a host computer at a fixed frequency. Table I presents the functions and DOFs of the servos.

The servos of XIN-REN, which are covered with silicone skin, act like human muscles to control robot's facial expressions. However, XIN-REN's expressions are not as rich as the humans because its servos and DOFs are fewer than those in humans. Therefore, the proposed system aims at enabling the humanoid robot to deliver expressions similar to those of the human performer and demonstrate movement smoothness, based on expression detail imitation. The following major issues are addressed in this study: 1) building the forward kinematics model for mapping a servo displacement space into an expression shape space and 2) establishing the inverse kinematics model, with the given facial expression series of the performer, to find the optimal servo displacements required for space–time similarity and movement smoothness.

IV. Forward Kinematics Modeling and Coefficient Matrix solving

Although facial expressions of a humanoid robot are directly controlled by adjusting its servo displacements, the movements of feature points can best embody facial expressions [1], [34]. Thus, the forward kinematics model is built between servo displacements and their corresponding feature points, and is

determined as an expression shape vector for a given servo displacement vector.

A. Forward Kinematics Model

To extract prominent facial features more precisely, 48 facial points, distributed on the mouth, eyebrows, eyes, and cheek, are tracked by the AAM algorithm because of its high efficiency and robust anti-interference [15], [35]. These points constitutes the expression shape vector S

$$S = (x_1, y_1, \ldots, x_i, y_i, \ldots, x_n, y_n)^T (n = 48)$$

where (x_i, y_i) represents the coordinates of the ith feature point. Because humanoid robots demonstrate facial expressions through servo displacements, the servo displacement vector C is composed of m servos

$$C = (c_1, \ldots, c_j, \ldots, c_m)^T (m = 7).$$

As the shape vector S depends only on the initial shape vector $S_0 = (x_{10}, y_{10}, \ldots, x_{i0}, y_{i0}, \ldots, x_{n0}, y_{n0})^T$, the initial servo displacement vector $C_0 = (c_{10}, \ldots, c_{j0}, \ldots, c_{m0})^T$ and the current servo displacement vector C, the forward kinematics model between C and S can be expressed as $\Psi(\bullet)$

$$S = \Psi(S_0, C_0, C). \tag{1}$$

However, a servo may influence multiple feature points and a feature point may be affected by multiple servos. The many-to-many nonlinear relationships between servos and feature points increase the complexity of solving $\Psi(\bullet)$[14, 26]. If m servos and n feature points are regarded as an autonomous loop system [36], the potential energy of each feature point can be transformed from the mechanical energy of the servos, based on the energy conservation principle [37]. According to this hypothesis, the horizontal potential energy $E_i^x = \mathrm{sgn}(x_i - x_{i0})(x_i - x_{i0})^2$ and vertical potential energy $E_i^y = \mathrm{sgn}(y_i - y_{i0})(y_i - y_{i0})^2$ of the feature point i can be expressed by the mechanical energy superposition of m servos ($\mathrm{sgn}(\bullet)$ is a symbolic function) as follows:

$$\begin{cases} E_i^x = a_{2i-11}E_1^C \cdots + a_{2i-1j}E_j^C \cdots + a_{2i-1m}E_m^C \\ \quad = \alpha_{2i-1}E^C \\ E_i^y = a_{2i1}E_1^C \cdots + a_{2\mathrm{ij}}E_j^C \cdots + a_{2\mathrm{im}}E_m^C = \alpha_{2\mathrm{i}}E^C \end{cases}. \tag{2}$$

In (2), $(E^C)_{m\times 1} = (E_1^C, ..., E_j^C, ..., E_m^C)^T$ represents the mechanical energy vector of m servos, $E_j^C = \mathrm{sgn}(c_j - c_{j0})(c_j - c_{j0})^2$ represents the mechanical energy of the jth servo, $\alpha_{2i-1} = (a_{2i-11}, \ldots, a_{2i-1j}, \ldots, a_{2i-1m})$ and $\alpha_{2i} = (a_{2i1}, \ldots, a_{2\mathrm{ij}}, \ldots, a_{2\mathrm{im}})$ represent the weight coefficients of m servos on E_i^x and E_i^y, respectively. For the potential energy of all feature points, the mapping relations can be expressed as

$$\begin{aligned} E^S &= (E_1^x, E_1^y, \ldots, E_i^x, E_i^y \ldots, E_n^x, E_n^y)^T \\ &= (\alpha_1, \alpha_2 \ldots, \alpha_{2i-1}, \alpha_{2i} \ldots, \alpha_{2n-1}, \alpha_{2n})^T E^C = AE^C \end{aligned} \tag{3}$$

where $(E^S)_{2n\times 1}$ and $(A)_{2n\times m}$ represent the potential energy of shape vector S and the coefficient matrix, respectively. As the horizontal and vertical potential energy of each feature point can be represented by a linear superposition of the mechanical energy of m servos in (2), the nonlinear model between the servo displacement vector C and its corresponding shape vector S in (1) is converted into the linear relationships between E^C and E^S in (3).

B. Coefficient Matrix Solution by Multiple Linear Regression

The weighting coefficients of each row in A reflect the strength of the mechanical energy of each servo on the potential energy of the corresponding feature point. However, obtaining an accurate expression for A is difficult because of hardware complexity. Therefore, a multiple linear regression method is adopted to solve for an approximate expression [38]. In solving for coefficient matrix A, K samples are collected, from which the following training set forms:

$$\begin{aligned} D = \big\{ & E^{C(k)} = (E_{1k}^C, \ldots, E_{\mathrm{mk}}^C), E^{S(k)} \\ & = (E_{1k}^x, E_{1k}^y, \ldots, E_{\mathrm{nk}}^x, E_{\mathrm{nk}}^y) \big\}_{k=1}^K. \end{aligned}$$

With K samples, the mechanical potential energy of m servos and the potential energy of n feature points can be represented by $(X)_{K\times m}$ and $(Y)_{K\times 2n}$, respectively,

$$X = \begin{pmatrix} E_{11}^C \cdots E_{j1}^C \cdots E_{m1}^C \\ \cdots \quad \cdots \quad \cdots \\ E_{1k}^C \cdots E_{\mathrm{jk}}^C \cdots E_{\mathrm{mk}}^C \\ \cdots \quad \cdots \quad \cdots \\ E_{1K}^C \cdots E_{\mathrm{jK}}^C \cdots E_{\mathrm{mK}}^C \end{pmatrix}$$

$$Y = \begin{pmatrix} E_{11}^x E_{11}^y \cdots E_{i1}^x E_{i1}^y \cdots E_{n1}^x E_{n1}^y \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ E_{1k}^x E_{1k}^y \ldots E_{\mathrm{ik}}^x E_{\mathrm{ik}}^y \ldots E_{\mathrm{nk}}^x E_{\mathrm{nk}}^y \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ E_{1K}^x E_{1K}^y \cdots E_{\mathrm{iK}}^x E_{\mathrm{iK}}^y \cdots E_{\mathrm{nK}}^x E_{\mathrm{nK}}^y \end{pmatrix}. \tag{4}$$

For the vertical weight coefficients of the ith feature point, solving for coefficient vector α_{2i} means minimizing the squared errors of E_i^y on training set D, and the objective function is formulated as follows:

$$\begin{aligned} \min_{a_{2i}} Q &= \sum_{k=1}^{K} (E_{\mathrm{ik}}^y - a_{2i1}E_{1k}^C \cdots - a_{2\mathrm{im}}E_{\mathrm{mk}}^C)^2 \\ &= (Y_i^y - X\alpha_{2i}^T)^T (Y_i^y - X\alpha_{2i}^T) \end{aligned} \tag{5}$$

where $(Y_i^y)_{K\times 1} = (E_{i1}^y, \cdots, E_{\mathrm{ik}}^y, \cdots, E_{\mathrm{iK}}^y)^T$ represents the vertical potential energy of the ith feature point with K samples. Consequently, (5) can be solved by the following equation:

$$\frac{\partial Q}{\partial \alpha_{2i}^T} = -2X^T(Y_i^y - X\alpha_{2i}^T) = 0.$$

When X^TX is reversible, the optimal value of a_{2i} can be expressed as follows:

$$\hat{\alpha}_{2i}^T = (X^TX)^{-1}X^TY_i^y. \tag{6}$$

For other coefficient vectors, the similar regression analysis is conducted on the training set D, and the optimal parameters

of coefficient matrix A are approximated by (6):

$$\begin{aligned} A^T &= (\hat{\alpha}_1^T, \hat{\alpha}_2^T, \ldots, \hat{\alpha}_{2i-1}^T, \hat{\alpha}_{2i}^T, \ldots, \hat{\alpha}_{2n-1}^T, \hat{\alpha}_{2n}^T) \\ &= (X^T X)^{-1} X^T (Y_1^x, Y_1^y, \ldots, Y_i^x, Y_i^y, \ldots, Y_n^x, Y_n^y) \\ &= (X^T X)^{-1} X^T Y. \end{aligned}$$

Thus,

$$A = (A^T)^T = ((X^T X)^{-1} X^T Y)^T = Y^T X (X^T X)^{-1}. \quad (7)$$

When the coefficients of A, which reflect the linear statistical relationships between the mechanical energy of the servos and the potential energy of the feature points, are obtained, the nonlinear forward kinematics model between C and S, can be decoupled by

$$\begin{aligned} S &= \Psi(S_0, C_0, C) = (\Psi_1(S_0, C_0, C), \ldots, \Psi_{2n}(S_0, C_0, C)) \\ &= S_0 + \left(\mathrm{sgn}(\alpha_1 E^C) \sqrt{|\alpha_1 E^C|}, \ldots, \mathrm{sgn}(\alpha_{2n} E^C) \right. \\ &\quad \left. \times \sqrt{|\alpha_{2n} E^C|} \right)^T. \end{aligned} \quad (8)$$

V. Inverse Kinematics Solution Based on Instantaneous Similarity and Movement Smoothness

The inverse kinematics model aims at solving the optimal servo displacements $C = (c_1, c_2, \ldots, c_m)$, given a target shape vector $S^* = (x_1^*, y_1^*, ..., x_n^*, y_n^*)$, based on the forward kinematics model [5]. As A is not a square matrix, direct matrix transform method is not suitable. Thus, optimization is required to obtain the optimal solution, and the target function can be defined as

$$\begin{aligned} \min_C \Delta = & \left\| S^* - \Psi(S_0, C_0, C) \right\| \\ & \left\| S^* - S_0 \left(\mathrm{sgn}\alpha_1 E^C \sqrt{|\alpha_1 E^C|}, \ldots, \right. \right. \\ & \left. \left. \mathrm{sgn}(\alpha_{2n} E^C) \sqrt{|\alpha_2 E^C|} \right) \right\| \\ & s.t.\ 0 \le c_j \le 255, j = 1, 2, \ldots, m. \end{aligned} \quad (9)$$

In (9), shape consistency is considered without movement smoothness. However, facial expression learning for a robot is a dynamic process; it should maintain not only the shape consistency with the target shape vector for instantaneous similarity, but also movement smoothness to avoid zigzag effect and to protect servo hardware and facial silicone skin [18], [39], [40]. In addition, the humans are more sensitive to movement smoothness during human–computer interactions [3]. Apart from the forward kinematics model for reflecting the nonlinear relationships between the servo displacement vectors and their corresponding shape vectors, servo movement trajectory also is addressed for movement smoothness. Thus, trajectory prediction model is established to wrap servo displacements based on a polynomial function.

The movement trajectories for the servo displacements $C(t) = (c_1(t), \ldots, c_j(t), \ldots, c_m(t))^T$ at moment t can be expressed by a set of orthogonal polynomial time functions

$$C(t) = \Phi(t) = \left(\sum_{p=0}^{P} b_p^1 t^p, \ldots, \sum_{p=0}^{P} b_p^j t^p, \ldots, \sum_{p=0}^{P} b_p^m t^p \right)^T \quad (10)$$

where P is the highest power of the polynomial function and $b^j = (b_0^j, \ldots, b_p^j \ldots, b_P^j)^T$ $(1 \le j \le m)$ represents the coefficient of polynomial function for the jth servo displacement. In solving for these coefficients, the history series $C(t_k) = (c_1(t_k), \ldots, c_j(t_k), \ldots, c_m(t_k))^T (k = 1, 2, \ldots, d)$ of d moments before moment t, are collected, and the least squared error [41] is defined as

$$\min_{b^j} Q_j = \sum_{k=1}^{d} \left(c_j(t_k) - \sum_{p=0}^{P} b_p^j {t_k}^p \right)^2, 1 \le j \le m.$$

By solving m objective functions, the estimate values of $\{b^j\}_{j=1}^m$ can be obtained as follows:

$$B = (b^1, \ldots, b^j, \ldots, b^m) = (R^T R)^{-1} R^T U \quad (11)$$

where

$$R = \begin{pmatrix} 1 t_1 & \cdots & t_1^{P-1} t_1^P \\ \cdots & \cdots & \cdots \\ 1 t_k & \cdots & t_k^{P-1} t_k^P \\ \cdots & \cdots & \cdots \\ 1 t_d & \cdots & t_d^{P-1} t_d^P \end{pmatrix}$$

$$U = \begin{pmatrix} c_1(t_1) \cdots c_j(t_1) \cdots c_m(t_1) \\ \cdots \\ c_1(t_k) \cdots c_j(t_k) \cdots c_m(t_k) \\ \cdots \\ c_1(t_d) \cdots c_j(t_d) \cdots c_m(t_d) \end{pmatrix}. \quad (12)$$

In (12), $t_k = t - k \bullet T$ $(1 \le k \le d)$ and T is the interval between adjacent frames. With the obtained coefficient values, the servo displacements at moment t can be predicted as

$$\begin{aligned} C(t) &= \left(\sum_{p=0}^{P} b_p^1 t^p, \ldots, \sum_{p=0}^{P} b_p^j t^p, \ldots, \sum_{p=0}^{P} b_p^m t^p \right)^T \\ &= (uB)^T = \left(\left(R^T R \right)^{-1} R^T U \right)^T u^T = U^T R \left(R^T R \right)^{-1} u^T \end{aligned} \quad (13)$$

where $u = (1, t, \ldots, t^{P-1}, t^P)$. Given the target shape vector $S^*(t)$ and the servo trajectory prediction $\Phi^*(t)$ at moment t, and by combining the forward kinematics model and the trajectory prediction model, the target function (9) can be improved, and the optimal servo displacement vector $C(t)$ can be obtained by using

$$\begin{aligned} \min_{C(t)} J(C(t)) = & g \left\| S^*(t) - \Psi(S_0, C_0, C(t)) \right\| \\ & + (1 - g) \left\| C(t) - \Phi^*(t) \right\| \\ s.t. \quad & 0 \le c_j(t) \le 255, j = 1, 2, \ldots, m. \end{aligned} \quad (14)$$

In (14), the first item allows for the least shape deviation for instantaneous similarity, and the second item ensures movement

smoothness under the constraints of the predicted servo trajectory. The factor $g \in [0,1]$ is used for weighting instantaneous similarity and movement smoothness. The higher the g value is, the more weight should be given to the instantaneous similarity; conversely, the smaller the g value is, more weight should be given to movement smoothness.

For solving the nonlinear optimization problem with multiple constraints in (14), the exterior point penalty function method [16] is used. The new penalty function can be expressed as

$$\min \varphi\left(C\left(t\right), r_k\right) = \begin{cases} J\left(C\left(t\right)\right) & \text{(in the feasible domain)} \\ J\left(C\left(t\right)\right) + r_k \left(\left(C\left(t\right)-255\right)^T \left(C\left(t\right)-255\right) \right. \\ \left. +C(t)^T C\left(t\right)\right) & \\ & \text{(outside the feasible domain)} \end{cases} \tag{15}$$

where $r_k (k \in [1, +\infty))$ is the penalty factor meeting the incremental sequences $1, 2, \ldots, 2^k, \ldots$. Thus, the multiconstraint optimization problem $J(C(t))$, with $2m$ conditions, is converted into unconstrained optimization problem $\varphi(C(t), r_k)$, which can be solved by a gradient descent algorithm, and its iterative formula is expressed as follows:

$$C_k^{q+1}(t) = C_k^q(t) - \gamma \nabla \varphi\left(C_k^q(t), r_k\right) \tag{16}$$

where γ is the constant learning rate, $C_k^q(t)$ represents the servo displacements of the qth step iteration when $r_k = 2^k$, and $\nabla\varphi(C(t), r_k)$ represents the gradient information of $C(t)$, which can be expressed as

$$\nabla \varphi\left(C(t), r_k\right) = \begin{cases} \dfrac{dJ\left(C(t)\right)}{dC(t)} & \text{(in the feasible domain)} \\ \dfrac{dJ\left(C(t)\right)}{dC(t)} + r_k\left(2C(t) - 255\right) & \\ & \text{(outside the feasible domain)} \end{cases} \tag{17}$$

where

$$\begin{aligned} \frac{dJ\left(C(t)\right)}{dC(t)} = & -2g \frac{d\Psi\left(S_0, C_0, C(t)\right)}{dC(t)} \\ & \times \left(S^*(t) - \Psi\left(S_0, C_0, C(t)\right)\right) \\ & - \left(1-g\right)\left(C(t) - \Phi^*(t)\right) \end{aligned}$$

and

$$\frac{d\Psi\left(S_0, C_0, C(t)\right)}{dC(t)} = \begin{pmatrix} \dfrac{d\Psi_1(\bullet)}{dc_1(t)} & \cdots & \dfrac{d\Psi_l(\bullet)}{dc_1(t)} & \cdots & \dfrac{d\Psi_{2n}(\bullet)}{dc_1(t)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \dfrac{d\Psi_1(\bullet)}{dc_j(t)} & \cdots & \dfrac{d\Psi_l(\bullet)}{dc_j(t)} & \cdots & \dfrac{d\Psi_{2n}(\bullet)}{dc_j(t)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \dfrac{d\Psi_1(\bullet)}{dc_m(t)} & \cdots & \dfrac{d\Psi_l(\bullet)}{dc_m(t)} & \cdots & \dfrac{d\Psi_{2n}(\bullet)}{dc_m(t)} \end{pmatrix}$$

Fig. 3. Experimental setup used for data collection for solving forward kinematics model.

where

$$\frac{d\Psi_l(\bullet)}{dc_j(t)} = \begin{cases} \dfrac{sgn(\alpha_{2i-1}E^C)sgn\left(c_j(t)-c_{j0}\right)a_{2i-1j}\left(c_j(t)-c_{j0}\right)}{2\sqrt{|\alpha_{2i-1}E^C|}} & \\ & (l = 2i-1) \\ \dfrac{sgn(\alpha_{2i}E^C)sgn\left(c_j(t)-c_{j0}\right)a_{2ij}\left(c_j(t)-c_{j0}\right)}{2\sqrt{|\alpha_{2i}E^C|}} & \\ & (l = 2i) \end{cases}.$$

The final optimal servo displacements $C^*(t)$ with movement similarity and smoothness at moment t are calculated as

$$C^*(t) = \arg\min_{C(t)} J\left(C(t)\right) = \arg\min_{C(t)} \varphi\left(C(t), r_{k+1}\right). \tag{18}$$

VI. Automatic Facial Expression Learning Method

Human imitation is a straightforward method for humanoid robots to achieve human-like facial expressions [1], [7]. If the target shape vector $S^*(t)$ originates from a human performer, the optimal servo displacements $C^*(t)$ can be calculated based on the inverse kinematics model so that the facial expressions of the robot are similar to those of the performer. Thus, an automatic expression learning method is proposed, which is divided into two stages: 1) the offline solution of the forward kinematics model and 2) the online facial expression learning from humans.

VII. Experiment and Analysis

A. Evaluation of the Proposed Automatic Expression Learning Method

1) Evaluation of the Forward Kinematics Model: We conduct experiments to assess the effectiveness and performance of the forward kinematics model, based on the energy conservation principle. Fig. 3 shows the experimental setup used for data collection. In the experiment setup, a Logitech camera with

Algorithm 1: Offline Solution of the Forward Kinematics Model.

Input: Initial servo displacement vector $C_0 = (c_{10}, \ldots, c_{j0} \ldots, c_{m0})$ and its corresponding initial shape vector $S_0 = (x_{10}, y_{10}, \ldots, x_{i0}, y_{i0}, \ldots, x_{n0}, y_{n0})$ of the robot, and size K of the training set $D = \{E^{C(k)}, E^{S(k)}\}_{k=1}^{K}$.

Output: Coefficient matrix A.

$k \leftarrow 1$.
while $k \leq K$ **do**
  $j \leftarrow 1$.
  while $j \leq m$ **do**
    $c_j \leftarrow 255 \times rand()$.
    $E_j^C \leftarrow sgn(c_j - c_{j0}) \times (c_j - c_{j0})^2$.
    $j \leftarrow j + 1$.
    end while
  $C \leftarrow (c_1, \ldots, c_j, \ldots, c_m)$.
  Drive the robot with the current servo displacements C.
  Gain current shape vector $S \leftarrow (x_1, y_1, \ldots, x_n, y_n)$.
  $i \leftarrow 1$.
  while $i \leq n$ **do**
    $E_i^x \leftarrow sgn(x_i - x_{i0}) \times (x_i - x_{i0})^2$.
    $E_i^y \leftarrow sgn(y_i - y_{i0}) \times (y_i - y_{i0})^2$.
    $i \leftarrow i + 1$.
    end while
  $E^{C(k)} \leftarrow (E_1^C, \ldots, E_j^C, \ldots, E_m^C)$,
  $E^{S(k)} \leftarrow (E_1^x, E_1^y, \ldots, E_i^x, E_i^y, \ldots, E_n^x, E_n^y)$.
  $k \leftarrow k + 1$.
  end while
  $X \leftarrow (E^{C(1)} \cdots E^{C(k)} \cdots E^{C(K)})^T$,
  $Y \leftarrow \left(E^{S(1)} \cdots E^{S(k)} \cdots E^{S(K)}\right)^T$.
  $A \leftarrow Y^T X (X^T X)^{-1}$.

Algorithm 2: Online Facial Expression Learning From Humans.

Input: Initial servo displacement vector C_0 and its corresponding initial shape vector S_0 of the robot, coefficient matrix A, polynomial power P, frame rate T, history movement window size d, weighting factor g, and termination threshold ε.

Output: Servo displacement series for the robot to generate facial expressions similar to those of the human performer.

$t \leftarrow 0$.
while true **do**
  Obtain the shape vector $S^*(t)$ of human performer from a camera at moment t.
  Align $S^*(t)$ with S_0 by singular value decomposition method [42].
  if $t <= d$ then
  $C^*(t) \leftarrow \arg\min_{C(t)} \|S^*(t) - \Psi(S_0, C_0, C(t))\|$.
  else
    $k \leftarrow 1$.
    while $k \leq d$ **do**
      $t_k \leftarrow t - k \times T$.
      $R_k \leftarrow (1 t_k \cdots t_k^{P-1} t_k^P)$.
      $U_k \leftarrow (c_1(t_k) \cdots c_j(t_k) \cdots c_m(t_k))$.
      $k \leftarrow k + 1$.
      end while
  Set
  $R \leftarrow (R_1 \cdots R_k \cdots R_d)^T$, $U \leftarrow (U_1 \cdots U_k \cdots U_d)^T$,
  $u \leftarrow (1, t, \ldots, t^{P-1}, t^P)$.
  $\Phi^*(t) \leftarrow U^T R (R^T R)^{-1} u^T$.
  $C^*(t) \leftarrow \arg\min_{C(t)} \begin{pmatrix} g\|S^*(t) - \Psi(S_0, C_0, C(t))\| \\ +(1-g)\|C(t) - \Phi^*(t)\| \end{pmatrix}$.
  end if
  Send the optimal servo displacements $C^*(t)$ to the robot through built-in RS232C interface.
  $t \leftarrow t+T$
end while

HD1080P, installed at an optimum distance (40–50 cm) from the robot, is oriented directly toward the head of the robot and used for capturing facial expression images at 10 frames/s. To maintain a constant ambient light and best tracking performance, two LED light sources are placed at the side of the robot. To collect training data, $m(m= 7)$ servo displacements, which constitute the servo displacement vector, are randomly generated and sent to XIN-REN's servo controller for corresponding facial expressions at the same frequency as that of the HD camera. Within every interval (100 ms), the servo displacements remain stationary, and the open source AAMlibrary-2.5 is applied to track its corresponding n ($n= 48$) feature points for the expression shape vector S (48×2 dimensions). To eliminate rigid head and shoulder motions, S is scaled and rotated to match the initial shape S_0. Synchronously, the pairs of C and its corresponding S, which is regarded as the ground truth for actual feature points positions, are recorded for solving the coefficient matrix A.

Using the experimental setup shown in Fig. 3, 15 000 facial images $\{(C(k), S(k))\}_{k=1}^{15\,000}$ are collected and used for verifying the rationality and reliability of the forward kinematics model. To ensure that the test samples will not be used during training, 15 000 normalized samples are randomly divided into two groups. The first group comprises the K samples, which are grouped as the training set $D = \{(E^{C(k)}, E^{S(k)})\}_{k=1}^{K=10\,000}$ and used for training the forward kinematics model and obtaining coefficient matrix A. The second group, comprising the remainder of the samples, functions as the testing set $T = \{(E^{C(k)}, E^{S(k)})\}_{k=1}^{q=500}$. The root mean squared error [7], [27] is used to evaluate the shape deviation $\mathrm{RMSE}_{S(k)}$ of the test sample k, and the position deviation RMSE_i of feature point i

$$\mathrm{RMSE}_{S(k)} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_{\mathrm{ik}} - \hat{x}_{\mathrm{ik}})^2 + (y_{\mathrm{ik}} - \hat{y}_{\mathrm{ik}})^2}$$

$$\mathrm{RMSE}_i = \frac{1}{q}\sum_{k=1}^{q}\sqrt{(x_{\mathrm{ik}} - \hat{x}_{\mathrm{ik}})^2 + (y_{\mathrm{ik}} - \hat{y}_{\mathrm{ik}})^2} \quad (19)$$

TABLE I
FUNCTIONS AND DOFS OF SEVEN HEAD SERVOS

Servos	Function	DOFs
CH1	Eyebrow up/down	2
CH2	Cheek upward	1
CH3	Eyelid opening/closing	2
CH4	Eyeball right/left	2
CH5	Eyeball up/down	2
CH6	Mouth opening/closing	2
CH7	Mouth corner contraction	1

TABLE II
RESULTS FROM THREE FORWARD KINEMATICS MODEL METHODS

Method	Shape deviation(pixels)	Position deviation(pixels)
linear model	9.12 ± 2.34	6.71 ± 1.13
BP model[33]	5.31 ± 1.23	3.25 ± 0.43
Our proposed model	3.56 ± 1.43	2.09 ± 0.95

where $(x_{\mathrm{ik}}, y_{\mathrm{ik}})$ and $(\hat{x}_{\mathrm{ik}}, \hat{y}_{\mathrm{ik}})$ represent the actual position tracked by the AAM algorithm and the output position estimated by the forward kinematics model of feature point i, respectively. To further explain the advantages of the proposed model, the proposed nonlinear forward kinematics model is compared with a linear model, in which the mapping relation is only fitted by a one-order linear equation, and a BP model [33]. Meanwhile, to further test the robustness of different models and reduce randomness, the average predictable deviations of ten experiments, which involve different and random partitioning of samples into the training and testing sets, are counted through ten-fold cross validation. The mean shape deviation and the mean position deviation of the three models are shown in **??** Table II. The results in Table II show that the mean shape deviation and the mean position deviation of the proposed model does not exceed 6 pixels and 3 pixels (with the face region $w = 100$, $h = 150$), respectively. Compared with other models, the shape and position deviations of the proposed model are smaller, whereas its variance is not as good as that of the BP model.

To further illustrate the significant differences between different models, the Student's t test method is applied to multiple pairwise comparison tests. Meanwhile, to counteract the problem of multiple comparisons, Bonferroni's multiple comparisons correction [43], which is regarded as the simplest and most conservative method to control the family-wise Type I error, is applied. Instead of following the procedure of adjusting the statistical significance level by dividing it with the number of comparisons, the equivalent procedure of adjusting the p-value of each hypothesis by multiplying it with the number of comparisons is adopted [44]. Specifically, the two independent hypotheses (linear model versus our proposed model, and BP model versus our proposed model) are tested for each deviation index (shape deviation or position deviation) with the same data at 0.05 significance level; thus, the adjusted p-value holds twice the p-value. Both these values, as well as 95% confidence intervals, are reported in Table III, which shows significant differences at the adjusted p-value < 0.01 between the

TABLE III
DIFFERENCES FORM MULTIPLE PAIRWISE COMPARISON TESTS

Method	Our proposed model			
	Deviation	p-value	Adjusted p-value	95% Confidence Interval for Mean
linear model	Shape deviation	<0.0001	<0.0001	[3.70 7.38]
	Position deviation	<0.0001	<0.0001	[3.68 5.57]
BP model[33]	Shape deviation	0.0114	0.0228	[0.50 3.00]
	Position deviation	0.0013	0.0026	[0.59 1.74]

scores for our proposed model and the linear model. Moreover, our proposed model is still superior to the BP model because of the adjusted p-value < 0.05, although the differences between the two models are not significant at the adjusted p-value>0.01. Considering the smaller deviations in the mean shape and the mean position of our proposed model (see Table II), as well as the adjusted p-values of the pairwise comparison tests (see Table III), we conclude that our proposed model has better reliability in shape deviation and position deviation than the linear model and the BP model.

The position deviation RMSE_i in (19) reflects the holistic displacement deviation of feature point i; however, our forward kinematics model is built based on horizontal potential energy and vertical potential energy. Hence, the normalized horizontal deviation RMSE_i^x and the normalized vertical deviation RMSE_i^y for feature point i are calculated as follows:

$$\mathrm{RMSE}_i^x = \frac{1}{q}\sum_{k=1}^{q}\sqrt{\left(\frac{x_{\mathrm{ik}} - \hat{x}_{\mathrm{ik}}}{w}\right)^2},$$

$$\mathrm{RMSE}_i^y = \frac{1}{q}\sum_{k=1}^{q}\sqrt{\left(\frac{y_{\mathrm{ik}} - \hat{y}_{\mathrm{ik}}}{h}\right)^2}.$$

Fig. 4 shows the statistical results, which indicate certain differences between the predicted deviations of 48 feature points. For instance, the deviations (RMSE_3^x and RMSE_3^y) of feature point 3 (the tip of nose) are zero because this point is used as the alignment reference point for each test sample. These are the greatest normalized horizontal deviations at $\mathrm{RMSE}_7^x = 2.3\%$ (just as 2.3 pixels with the width of face region $w = 100$), and the greatest normalized vertical deviations $\mathrm{RMSE}_{12}^y = 2.4\%$ (just as 3.6 pixels with the height of face region $h = 150$).

As the feature points are distributed mainly in the facial regions, such as eyebrow, eyes, and mouth, the deviations of these facial regions can very well reflect the prediction ability of forward kinematics model. From the deviations statistics of six facial regions, presented in Table IV, it can be seen that lower horizontal deviations are in the mouth and eyeball regions, and better prediction performance for vertical deviations in the cheek and eyes regions, whereas larger horizontal and vertical deviations are in the jaw region. A reasonable explanation for these observations is that the facial regions (such as eyes, mouth, cheek) are influenced by only a single servo, whereas the jaw region is codetermined by multiple servos owing to kinematic constraints.

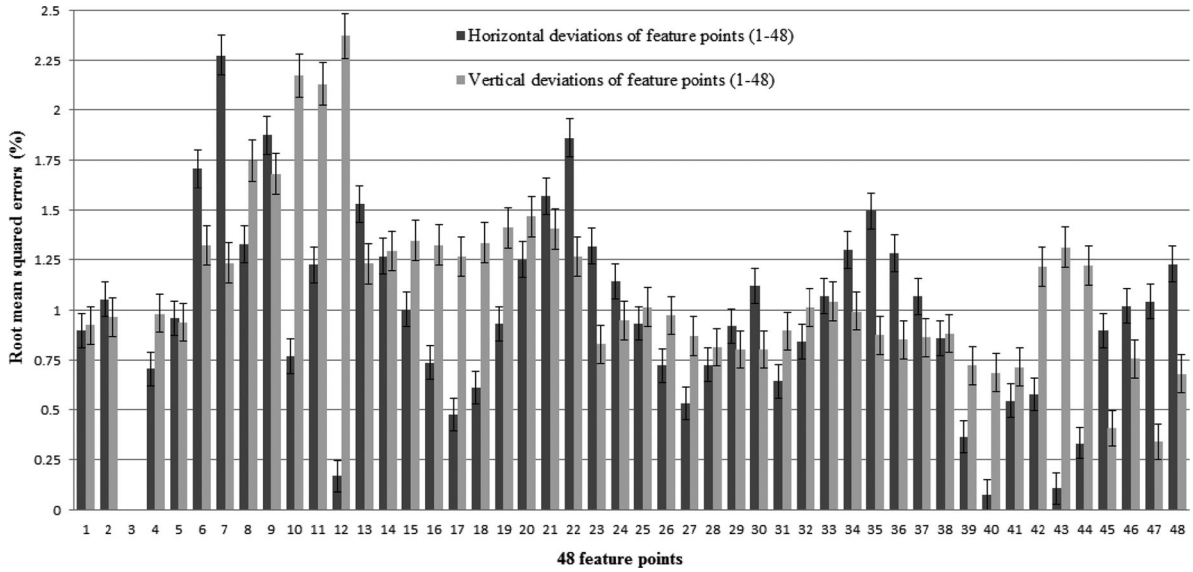


Fig. 4. Horizontal and vertical deviations for feature points 1-48.

TABLE IV
COMPARISON OF DEVIATIONS FOR DIFFERENT FACIAL REGIONS

Facial regions	Serial of feature points	Number of feature points	Horizontal deviations(%)	Vertical deviations(%)
Eyebrow	13– 22	10	1.12 ± 0.45	1.33 ± 0.07
Eyes	23–38	16	1.00 ± 0.27	0.90 ± 0.08
Eyeball	1-2	2	0.97 ± 0.11	0.94 ± 0.03
Mouth	4, 5, 39–44	8	0.46 ± 0.30	0.97 ± 0.25
Cheek	45–48	4	1.05 ± 0.14	0.55 ± 0.12
Jaw	6–12	7	1.33 ± 0.71	1.81 ± 0.44

The foregoing evaluation results show that our proposed forward kinematics model, which is built based on the energy conservation principle, not only has littler shape deviations and position deviations, but also retains smaller deviation in key facial regions, such as the mouth and the eyes, which play a critical role in characterizing different expressions.

2) Evaluation of the Inverse Kinematics Solution Based on Movement Similarity and Movement Smoothness: Inverse kinematics solution is the process of finding the optimal servo displacements to maintain movement similarity with a human performer and movement smoothness with servo trajectory. Differing from the experiment setup for the forward kinematics model evaluation in which the data is recorded when the robot is in random state, the data for the inverse kinematics solution evaluation is collected by driving the robot to follow fixed sequences designed by animators in advance.

First, six expression categories (happy, anger, disgust, sad, surprise, fear) of the robot are arranged by animators, and five facial sequences with neutral-peak-neutral facial expression intensity variations are collected for each category. For each sequence, the shape vector $S(t)$ at moment t, and the $d(d = 10)$ history servo displacement vectors $\{C(t_i)\}_{i=1}^{d}$ before moment t, are used as inputs, and the servo displacement vector $C(t)$ is regarded as the output. Then, $M(M = 500)$ testing vectors at different moments t_k are randomly collected and grouped as

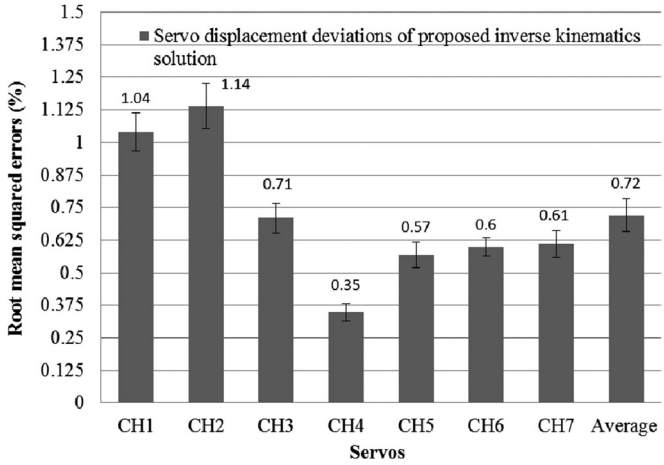


Fig. 5. Statistics from servo displacement deviations with $g = 0.6$.

$T1 = \{((S(t_k), \Phi(t_k)), C(t_k))\}_{k=1}^{M}$ for the inverse kinematics solution ($\varepsilon = 10^{-4}$); the average servo deviation RMSE_j^C for servo j can be defined as

$$\mathrm{RMSE}_j^C = \frac{1}{M}\sum_{k=1}^{M}\sqrt{\left(\frac{\hat{c}_{\mathrm{jk}} - c_{\mathrm{jk}}}{255}\right)^2}, j = 1, 2, \ldots, 7$$

where c_{jk} and $\hat{c}_{\mathrm{jk}}$ represent the actual displacement designed by animators and the output displacement estimated by inverse kinematics solution of servo j, respectively. The performance of inverse kinematics solution based on movement similarity is tested by using a ten-fold cross-validation. Fig. 5 shows the servo displacement deviations for M test samples with $g = 0.6$, which will be discussed in the forthcoming section on influence of parameter setting.

Fig. 5 shows that the servo displacement deviations, CH4 (Eyeball right/left) and CH5 (Eyeball up/down), with relatively fewer feature points, are small. Ideally, the smaller the servo displacement deviations, the better the instantaneous similarity

with target servo displacements. In practice, the displacement deviations for servos, associated with mouth (CH6), eyelids (CH3), and eyebrows (CH1), are relatively large. This can be attributed to the large number of feature points influenced by these servos, which results in the inaccuracy of the reverse solution of the servo displacements. But, viewed against the average servo displacement deviations, which do not exceed 0.8% (just as 2 with the maximum of servo displacement 255), the inverse kinematics solution can still be considered as having good predictive ability for finding the optimal servo displacements to maintain movement similarity.

Although Fig. 5 reflects the performance of the inverse kinematics solution with an instantaneous similarity, it measures neither the movement similarity nor smoothness of the imitation trajectory. Since these sequential indicators are more important for the subjective experience of humans [8], we further evaluate the performance of an online expression learning from a human performer, who performs all kinds of facial expressions with neutral-peak-neutral variations. The expression shape vectors of the performer are captured at the rate of 30 frames/s and regarded as the robot's target expression shape vectors. The sequential indicators of space-similarity G^S, time-similarity G^T, and movement smoothness G^D, measured by servo hopping during t_1 to t_L, following the method proposed by Zhu *et al.* [45], are defined as

$$\begin{cases} G^S = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{L}\sum_{k=1}^{L} F\left(d_i^H(t_k) - d_i^R(t_k), b_S\right)\right) \\ G^T = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{L}\sum_{k=1}^{L} F\left(d_i^H(t_k) - d_i^H(t_{k-1}) \right.\right. \\ \qquad \left.\left. - \left(d_i^R(t_k) - d_i^R(t_{k-1})\right), b_T\right)\right) \\ G^D = 1 - \frac{1}{L}\sum_{k=1}^{L}\frac{1}{m}\sum_{j=1}^{m} G\left(c_j(t_k)\right) \end{cases}$$

where $L = T = 30$ is the frame rate of camera;$(x_i^H(t), y_i^H(t))$ and $(x_i^R(t), y_i^R(t))$ represent the ith feature point positions of the performer and robot at moment t, respectively; (x_{i0}^H, y_{i0}^H) and (x_{i0}^R, y_{i0}^R) represent the ith feature point initial positions of the performer and robot, respectively; $d_i^H(t) = \sqrt{\left(x_i^H(t) - x_{i0}^H\right)^2 + \left(y_i^H(t_k) - y_{i0}^H\right)^2}$, $d_i^R(t) = \sqrt{\left(x_i^R(t) - x_{i0}^R\right)^2 + \left(y_i^R(t_k) - y_{i0}^R\right)^2}$ are the ith feature point displacement of the performer and robot at moment t, respectively; $F(x, b) = e^{-x^2/b}$ is a fitting function that converts the deviation parameter x to 0–1 similarity, and b is the parameter used to control the mapping performance. $G(c_j(t_k))$ indicates whether the displacement of jth servo exists unsmoothed hopping at moment t_k, and is measured as

$$G(c_j(t_k)) = \begin{cases} 1, & ||c_j(t_k) - c_j(t_{k-1})| - |c_j(t_{k-1}) - c_j(t_{k-2})|| > T_D \\ 0, & \text{other.} \end{cases}$$

Eventually, through ten-fold cross-validation, b_S is set to 0.3 for space similarity, b_T is set to 0.5 for time similarity, and T_D is set to 10. The statistical results, based on 20 expression series learned from the performer, illustrate that the average space similarity, time similarity, and movement smoothness are 93.1, 90.5, and 86.8, respectively. The values of average space similarity and time similarity reflect the similarity of the imitation trajectory with the performer's facial actions, whereas the value of the average movement smoothness reflects the smoothness of continuous servo motions.

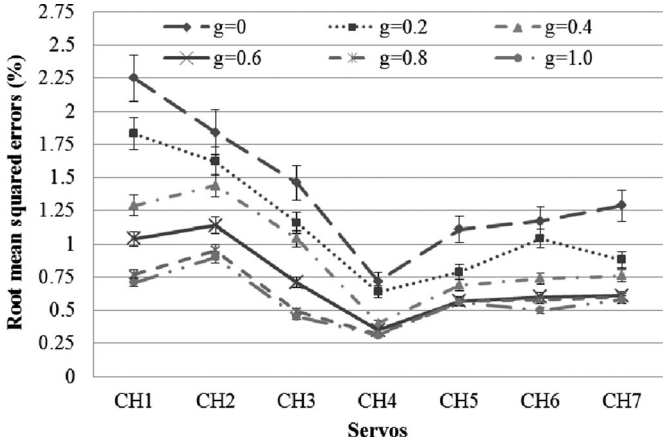


Fig. 6. Statistics of servo displacements with different g values.

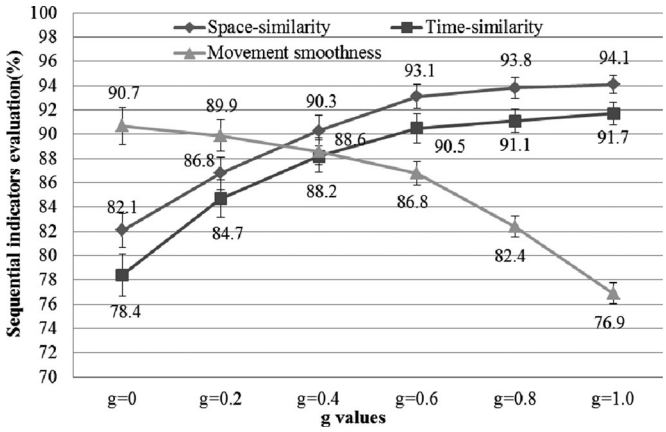


Fig. 7. Statistics of three sequential indicators with different g values.

3) Influence of Parameter g Setting: The effects of parameter setting of weighting factor g are discussed in this section. The parameter g is introduced in (14) to weight instantaneous similarity and movement smoothness. To investigate its effect on servo displacements and three sequential indicators, we change g from 0 to 1 by a step of 0.2, and the experimental results are shown in Figs. 6 and 7. Fig. 6 illustrates that the servos have large displacement deviations when g is zero ((14) considers only movement smoothness, but not movement similarity), and the displacement deviations gradually drop with increase in g. However, Fig. 7 shows, with increase in g, the space similarity and time similarity increase, whereas the movement smoothness descends. Hence, g is set to 0.6 for balancing movement similarity and movement smoothness.

4) Evaluation of the Different Methods: The purpose of this experiment is to evaluate the robot expressions generated using

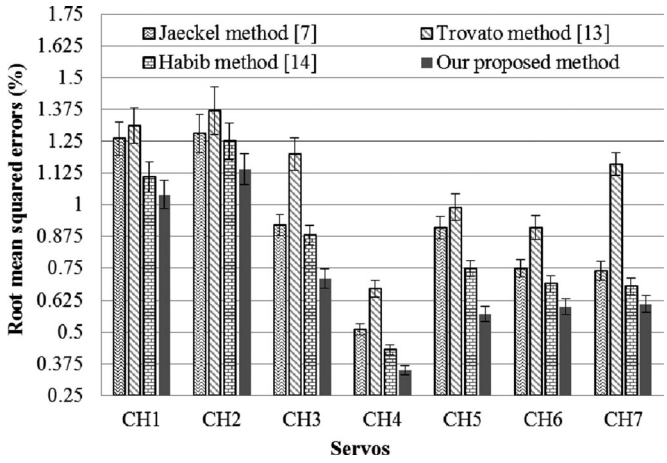


Fig. 8. Comparison of servo displacement deviations versus 4 methods.

TABLE V
COMPARISON OF SEQUENTIAL INDICATORS VERSUS FOUR METHODS

Sequential indicators	Space-similarity	Time-similarity	Movement smoothness
Jaeckel method [7]	85.4	85.2	83.3
Trovato method [13]	84.4	83.9	87.6
Habib method [14]	87.8	86.1	84.3
Our proposed method	93.9	91.5	86.1

different methods. To this end, the XIN-REN servo displacement deviations are compared with three sequential indicators using the state-of-the-art humanoid expression generation systems (Jaeckel method [7], Trovato method [13], and Habib method [14]). Based on the same testing set $T1$, the prediction deviations of different methods are compared. Fig. 8 indicates that the four methods present small prediction deviations for eyeball movement (CH4, CH5) and the opening/closing of the mouth (CH6), but large prediction deviations for the upward/downward movement of the eyebrows (CH1) and upward movement of the cheek (CH2). However, our proposed method presents smaller prediction deviations than other methods, thus establishing the validity of the inverse kinematics solution, based on the constraints of movement similarity and movement smoothness.

Meanwhile, 20 expression series from the performer are captured for the robot for online imitation. The performances of the methods in terms of three sequential indicators are shown in Table V. The results show that the Trovato method [13], which employs a facial cue model and polynomial classifier to solve the robotic cue variables from a prepared facial cue table, presents lower movement similarity but higher movement smoothness, because the facial cues are checked by animators. The movement similarity and the smoothness of the linear Jaeckel method [7], which is based on position information, are inferior to the proposed nonlinear method and the Habib method [14], which builds the mapping relations based on a neural network. Compared with the Habib method [14], the proposed method not only exhibits accurate servo displacements for a single-frame instantaneous similarity but also retains smooth motion for the multiframe expression imitation under the constraints of predicted servo trajectory.

B. Limitation of Our Method

The evaluations conducted prove that the proposed system can reproduce a natural and less hardwired robot expression. However, the proposed method has some limitations.

1) A person-specific AAM [46] instead of a generic AAM is used to fit the face model to improve the precision of 48 feature points [47]. However, the person-specific AAM is designed to model the appearance variation of a single person only. Thus, the online expression learning method demonstrates that the imitation performance of performers, whose images are not used for training AAM model, is poor.
2) The solution of the nonlinear forward kinematics model, which is converted into the linear relationships based on the energy conservation principle, weakens the many-to-many relationships between the servos and the feature point positions to some degree, as a result of which the deviations of servos, associated with more feature points, become larger.
3) During the experiment, some subtle facial features, such as squint, mouth contraction, frown, and upward cheek movement, demonstrate small shape deviations with respect to the target shape vectors of the performer, even when no actual mapping quality is maintained, in terms of human visual characteristics.

VIII. CONCLUSION

Subject to the complex kinematic constraints of the humanoid robot XIN-REN, an automatic facial expression learning method has been proposed by extending its facial expressions through learning from a human performer. The evaluation results establish the rationality and reliability of the forward kinematics model and the validity of the inverse kinematics solution under the constraints of instantaneous similarity and movement smoothness. In addition, the influence of g on the sequential indicators is discussed. Compared with other state-of-the-art humanoid expression generation systems, the proposed system not only keeps lower servo displacement deviations for a single frame instantaneous similarity, but also retains better movement smoothness for multiple frames expression imitation.

ACKNOWLEDGMENT

The authors would like to thank T. Cootes for providing calibration tools and Y. Wei for providing AAMlibrary-2.5.

REFERENCES

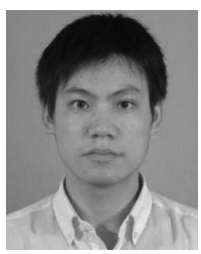
[1] V. Manohar and J. W. Crandall, "Programming robots to express emotions: Interaction paradigms, communication modalities, and context," *IEEE Trans. Human-Mach. Syst*, vol. 44, no. 3, pp. 362–373, Jun. 2014.
[2] J. W. Park, H. S. Lee, and M. J. Chung, "Generation of realistic robot facial expressions for human robot interaction," *J. Intell. Robot. Syst.*, vol. 78, no. 3, pp. 443–462, Jun. 2015.
[3] H. S. Ahn, D. W. Lee, D. Choi, D. Y. Lee, H. G. Lee, and M. H. Baeg, "Development of an incarnate announcing robot system using emotional interaction with humans," *Int. J. Humanoid Robot.*, vol. 10, no. 2, pp. 1–24, Jun. 2013.

[4] D. C. Lin, D. Godbout, and A. N. Vasavada, "Assessing the perception of human-like mechanical impedance for robotic systems," *IEEE Trans. Human-Mach. Syst*, vol. 43, no. 5, pp. 479–486, Sep. 2013.
[5] A. Dhall and R. Goecke, "Facial performance transfer via deformable models and parametric correspondence," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 9, pp. 1511–1519, Sep. 2012.
[6] G. Gibert, Y. Leung, and C. J. Stevens, "Control of speech-related facial movements of an avatar from video," *Speech Commun.*, vol. 55, no. 1, pp. 135–146, Jan. 2013.
[7] P. Jaeckel, N. Campbell, and C. Melhuish, "Facial behavior mapping-From video footage to a robot head," *Robot. Auton. Syst.*, vol. 56, no. 12, pp. 1042–1049, Dec. 2008.
[8] H. Yu and H. H. Liu, "Regression-based facial expression optimization," *IEEE Trans. Human-Mach. Syst*, vol. 44, no. 3, pp. 386–394, Jun. 2014.
[9] Y. Tadesse and S. Priya, "Graphical facial expression analysis and design method: An approach to determine humanoid skin deformation," *J. Mech. Robot.*, vol. 4, pp. 021010-1–021010-16, May 2012.
[10] H. Kamide, Y. Mae, T. Takubo, K. Ohara, and T. Arai, "Direct comparison of psychological evaluation between virtual and real humanoids: Personal space and subjective impressions," *Int. J. Human-Comput. Stud.*, vol. 72, no. 5, pp. 451–459, May 2014.
[11] M. Shayganfar, C. Rich, and C. Sidner, "A design methodology for expressing emotion on robot faces," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots. Syst.*, Oct. 2012, pp. 4577–4583.
[12] Y. Kondo, K. Takemura, J. Takamatsu, and T. Ogassawara, "A gesture-centric android system for multi-Party human-robot interaction," *J. Human–Robot Interact.*, vol. 2, no. 1, pp. 133–151, Mar. 2013.
[13] G. Trovato, M. Zecca, T. Kishi, and N. Endo, "Generation of humanoid robot's facial expressions for context-aware communication," *Int. J. Humanoid Robot.*, vol. 10, no. 1, pp. 1350013-1–1350013-22, Apr. 2013.
[14] C. Becker-Asano and H. Ishiguro, "Evaluating facial displays of emotion for the android robot Geminoid F," in *Proc. IEEE Workshop Affective Comput. Intell.*, Apr. 2011, pp. 1–8.
[15] N. Smolyanskiy, C. Huitema, L. Liang, and S. E. Anderson, "Real-time 3D face tracking based on active appearance model constrained by depth data," *Image Vis. Comput.*, vol. 32, no. 11, pp. 860–869, Nov. 2014.
[16] M. H. Farag, W. A. Hashem, and H. H. Saleh, "A penalty function approach for solving inequality constrained optimization problems," *Int. J. Math. Archive*, vol. 5, no. 7, pp. 33–40, Jul. 2014.
[17] F. Wilbers, C. Ishi, and H. Ishiguro, "A blendshape model for mapping facial motions to an android," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2007, pp. 542–547.
[18] N. Mavridis, "A review of verbal and non-verbal human-robot interactive communication," *Robot. Auton. Syst.*, vol. 63, no. 1, pp. 22–35, Jan. 2015.
[19] F. REN and K. Matsumoto, "Semi-automatic creation of youth slang corpus and its application to affective computing," *IEEE Trans. Affective Comput.*, vol. 7, no. 2, pp. 176–189, Jul. 2015, doi: 10.1109/TAFFC.2015.2457915.
[20] A. Habib, S. K. Das, I. C. Bogdan, D. Hanson, and D. O. Popa, "Learning human-like facial expressions for android Phillip K. Dick," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, Aug. 2014, pp. 1159–1162.
[21] L. Canamero and J. Fredslund, "I show you how I like you—Can you read it in my face?" *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 31, no. 5, pp. 454–459, Sep. 2001.
[22] M. Taïx, M. T. Tran, P. Souéres, and E. Guigon, "Generating human-like reaching movements with a humanoid robot: A computational approach," *J. Comput. Sci.*, vol. 4, no. 4, pp. 269–284, Jul. 2013.
[23] P. M. Yanik, J. Manganelli, J. Merino, A. L. Threatt, J. O. Brooks, and K. E. Green, "A gesture learning interface for simulated robot path shaping with a human teacher," *IEEE Trans. Human-Mach. Syst*, vol. 44, no. 1, pp. 41–54, Feb. 2014.
[24] S. Ontañón, J. L. Montaña, and A. J. Gonzalez, "A dynamic-bayesian network framework for modeling and evaluating learning from observation," *Expert Syst. Appl.*, vol. 41, no. 11, pp. 5212–5226, Sep. 2014.
[25] T. Shimizu, R. Saegusa, S. Ikemoto, H. Ishiguro, and G. Metta, "Robust sensorimotor representation to physical interaction changes in humanoid motion learning," *IEEE Trans. Neural Netw. Learn. Syst*, vol. 26, no. 5, pp. 1035–1047, May 2015.
[26] H. S. Ahn, D. W. Lee, D. Choi, and D. Y. Lee, "Designing of android head system by applying facial muscle mechanism of humans," in *Proc. 12th IEEE-RAS Int. Conf. Humanoid Robot.*, Nov. 2012, pp. 799–804.
[27] H. S. Ahn, D. W. Lee, and B. Macdonald, "Development of a human-like narrator robot system in EXPO," in *Proc. 6th IEEE Int. Symp. Autom. Mechatronics*, Nov. 2013, pp. 7–12.
[28] Y. Tadesse, "Actuation technologies for humanoid robots with facial expressions (HRwFE)," *Trans. Control Mech. Syst.*, vol. 2, no. 7, pp. 337–349, Jul. 2013.
[29] S. Nefti-Meziani, U. Manzoor, S. Davis, and S. K. Pupala, "3D perception from binocular vision for a low cost humanoid robot NAO," *Robot. Auton. Syst.*, vol. 68, pp. 129–139, Jun. 2015.
[30] J. K. Yoo, B. J. Lee, and J. H. Kim, "Recent progress and development of the humanoid robot HanSaRam," *Robot. Auton. Syst.*, vol. 57, no. 10, pp. 973–981, Oct. 2009.
[31] C. Breazeal, *Designing Sociable Robots*. Cambridge, MA, USA: MIT Press, 2002.
[32] H. Ishiguro, "Scientific issues concerning androids," *Int. J. Robot. Res.*, vol. 26, no. 1, pp. 105–117, Jan. 2007.
[33] E. Magtanong, A. Yamaguchi, K. Takemura, J. Takamatsu, and T. Ogasawara, "Inverse kinematics solver for android faces with elastic skin," in *Latest Advances in Robot Kinematics*. Amsterdam, The Netherlands: Springer, Jun. 2012, pp. 181–188.
[34] T. Kanda, T. Miyashita, T. Osada, and Y. Haikawa, "Analysis of humanoid appearances in human–robot Interaction," *IEEE Trans. Robot.*, vol. 24, no. 3, pp. 725–735, Jun. 2008.
[35] F. Dornaika, A. Moujahid, and B. Raducanu, "Facial expression recognition using tracked facial actions: Classifier performance analysis," *Eng. Appl. Artif. Intell.*, vol. 26, no. 1, pp. 467–477, Jan. 2013.
[36] Y. T. Feng, K. Han, and D. R. J. Owen, "Energy-conserving contact interaction models for arbitrarily shaped discrete elements," *Comput. Methods Appl. Mech. Eng.*, vols. 205–208, no. 15, pp. 169–177, Jan. 2012.
[37] K. Sasaki and R. R. Neptune, "Muscle mechanical work and elastic energy utilization during walking and running near the preferred gait transition speed," *Gait Posture*, vol. 23, no. 3, pp. 383–390, Apr. 2006.
[38] H. Wang, L. Shangguan, J. Wu, and R. Guan, "Multiple linear regression modeling for compositional data," *Neurocomputing*, vol. 122, no. 25, pp. 490–500, Dec. 2013.
[39] T. Shimizu, R. Saegusa, S. Ikemoto, H. Ishiguro, and G. Metta, "Robust sensorimotor representation to physical interaction changes in humanoid motion learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 5, pp. 1035–1047, May 2015.
[40] K. Noda, H. Arie, Y. Suga, and T. Ogata, "Multimodal integration learning of robot behavior using deep neural networks," *Robot. Auton. Syst.*, vol. 62, no. 6, pp. 721–736, Jun. 2014.
[41] R. Montoliu and F. Pla, "Generalized least squares-based parametric motion estimation," *Comput. Vis. Image Underst.*, vol. 113, no. 7, pp. 790–801, Jul. 2009.
[42] W. Kim, S. Suh, W. Hwang, and J. J. Han, "SVD face: illumination-invariant face representation," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1336–1340, Nov. 2014.
[43] W. R Rice, "Analyzing tables of statistical tests," *Evolution*, vol. 43, no. 1, pp. 223–225, Jan. 1989.
[44] Y. Mohammad and T. Nishida, "Why should we imitate robots? Effect of back imitation on judgment of imitative skill," *Int. J. Soc. Robot.*, vol. 7, no. 4, pp. 497–512, Aug. 2015.
[45] T. Zhu, Q. Zhao, and Z. Xia. "A visual perception algorithm for human motion by a Kinect," *J. Robot.*, vol. 36, no. 6, pp. 647–653, Nov. 2014.
[46] R. Gross, I. Matthews, and S. Baker, "Generic vs. person specific active appearance models," *Image Vis. Comput.*, vol. 23, no. 11, pp. 1080–1093, Nov. 2005.
[47] F. Ren and Z. Huang, "Facial expression recognition based on AAM–SIFT and adaptive regional weighting," *IEEJ Trans. Electric. Electron. Eng.*, vol. 10, no. 6, pp. 713–722, Sep. 2015.

Fuji Ren (SM'03) received the Ph.D. degree in 1991 from the Faculty of Engineering, Hokkaido University, Sapporo, Japan.

He became a Professor in the Faculty of Engineering, University of Tokushima, Tokushima, Japan, in 2001. His research interests include artificial intelligence, language understanding and communication, and affective computing.

Prof. Ren is a Fellow of the Japan Federation of Engineering Societies.



Zhong Huang received the bachelor's degree in 2005 from Anqing Normal University, Anqing, China, and the master's degree, in 2008 from Hefei University of Technology, Hefei, China, where he is currently working toward the Ph.D. degree.

His research interests include humanoid robots, affective computing, and computer vision.