# A Facial Expression Imitation System in Human Robot Interaction

S. S. Ge, C. Wang, C. C. Hang

*Abstract*— In this paper, we propose an interactive system for reconstructing human facial expression. In the system, a nonlinear mass-spring model is employed to simulate twenty two facial muscles' tensions during facial expressions, and then the elastic forces of these tensions are grouped into a vector which is used as the input for facial expression recognition. The experimental results show that the nonlinear facial mass-spring model coupled with the SVM classifier is effective to recognize the facial expressions. Finally, we introduce our robot that can make artificial facial expressions. Experimental results of facial expression generation demonstrate that our robot can imitate six types of facial expressions.

### I. INTRODUCTION

As social robots become more and more interactive and communicated, it is crucial that they can understand, perceive and imitate the human emotions appropriately in the social environment [1]. As discussed in [2], synthesizing the aspects of empathy within robots is the first step to improve the robots' ability to engage in fruitful interactions with the human users. However, it is insufficient for empathetic interactions between humans and social robots only with the human face detection and recognition. Facial expression is the most effective nonverbal way for human beings to express their emotions and interact with others. Therefore, it is necessary for a social robot to recognize and imitate human facial expressions [3].

Within the recent decade, many researchers have been trying to automatically recognize facial expressions of human beings. Various pattern recognition methods have been used in order to recognize facial expressions. Facial expression recognition approaches could be divided into two main categories: target oriented and gesture oriented. Target oriented approaches [4], [5] attempt to infer the facial expression only from one typical facial image. Gesture oriented approaches [6] utilize temporal information from a sequence of facial expression motion images. As discussed in [7], [8], the effective use of the dynamic information of facial motions is critical to the emotion recognition and interpretation. The representative research work based on facial motions include [9], [10] and [11]. For example, a 3D face mesh is employed based on FACS model in [9]. The position of attachment and the elastic properties of the facial muscles are estimated in model-based facial image coding for facial expression recognition. In essence the 3D face mesh is simulated by a set of mass-springs. That is to say, each edge of the mesh can be simulated by a linear spring model.

Another area of research on human robot interaction (HRI) is to generate facial expressions of the robots. Currently, one of representative HRI systems is the mechanical looking robot. With respect to the mechanical looking robot, we must consider the following well-developed robotic faces. In 2004, researchers at Takanishi's laboratory developed a robot called the Waseda Eye No.4 or WE-4, which can communicate naturally with humans by expressing human-like emotions [12]. Before developing Leonardo, Breazeal's research group at the MIT developed an expressive anthropomorphic robot called Kismet, which engages people in natural and expressive face-to-face interaction. With 15 DOFs, the face of the robot can display a wide assortment of facial expressions which can reflect its emotional state [13].

However, the existing 3D face mesh for facial expression recognition is based on the assumption of linear mass-spring model. As discussed in [14], the simple linear mass-spring models can not simulate the real issue muscles accurately. Thus we employ nonlinear mass-spring model to simulate twenty two facial muscles' tensions for facial expression recognition. Given the recognition results, our robot can imitate six types of human facial expressions through motor actions, including happiness, surprise, sadness, disgust, fear and anger.

The remainder of this paper is organized as follows: In section 2, we illustrate the system used for facial expression recognition and imitation. In Section 3, we present the facial muscles and principle of facial expressions. In Section 4, we discuss the nonlinear mass-spring model which can be used to simulate the muscle's tension during the expression. In Section 5, we present how to classify the facial expressions and summarize the experimental results. Section 6 describes the proposed human-robot interaction application. Finally, we give some conclusions and discuss our future work.

### **II. SYSTEM DESCRIPTION**

As shown in Fig. 1, our goal is to build a system to imitate the human facial expressions. The proposed system is composed of four key modules: face detection, feature extraction, classification and artificial emotion generation.

The system framework is shown in Fig. 2. First the face detection module segments the face regions of a video sequence or an image and locates the positions of the eyebrows, eyes, nose and mouth. The positions can be represented by some driven points with special mathematic properties (i.e., the minima). The module of feature extraction is used to track the driven points during a facial expression, and compute their sequential displacements compared to their corresponding fixed points. In the system a facial muscle

S. S. Ge, C. Wang and C. C. Hang are with Social Robotics Lab, Interactive Digital Media Institute and Edutainment Robotics Lab, Department of Electrical and Computer Engineering, National University of Singapore, Singapore, {samge,g0600196,elehcc}@nus.edu.sg



Fig. 1: Robot imitates human facial expression.

is assumed to consist of a pair of key points, namely driven point and fixed point. The fixed points, which are derived from the facial mass-spring model, can not be moved during a facial expression. Given the outputs of feature extraction and a predefined set of facial expressions, the classification module classifies a video or an image into the corresponding class of facial expressions (i.e., happiness, fear, etc). Finally, the module of artificial emotion generation can control a social robot to imitate the facial expression in response of the user's expression.



Fig. 2: Overview of the system framework.

## III. FACE DETECTION AND KEY POINT EXTRACTION

The face detection is the preparatory step in the facial expression recognition. For example, it can fix a range of interests, decrease the searching range and initial approximation area for the feature selection. In the system, we assume that there is only one face in an image [3]. The face should take up a significant area in the image. OpenCV is a powerful tool for the task of face detection, therefore we employ OpenCV to segment the face region from an image and extract facial features [15].

Once the face region is segmented, facial feature extraction is conducted to locate the positions of the eyebrows, eyes, eyelids, mouth and wrinkles. The positions can be represented by some driven points with special mathematic properties (i.e., the minima).



Fig. 3: The facial mass-spring model

As discussed in [16] and Facial Action Coding System (FACS), there are twenty two facial muscles which are closely relevant to human expressions. In fact, the human facial expressions originate from the movements of facial muscles beneath the skin. Thus we represent each facial muscle by a pair of key points, namely driven point and fixed point. As shown in Figure 3, the driven points are marked by red points which can be moved during a facial expression, while the fixed points are remarked by blue points which can not be moved during a facial expression. Further, each yellow line represents a facial muscle. According to the facial muscle model [14], we can find the locations of these fixed points in the facial surface. From the viewpoint of facial muscle model, a facial expression can be regarded as a combination of the different movements of those driven points in different facial expressions. For example, we show the movements of driven points in Fig. 4. In the next section, we will discuss the problem of how to describe such movements by using mass-spring model.

# IV. FEATURE EXTRACTION AND FACIAL EXPRESSION RECOGNITION

Mass-spring model is typically utilized to formulate the facial muscle deformation [17]. That is to say, the facial muscle can be contracted or stretched like a spring. As usual, the facial muscle is treated as a linear spring and the elastic stiffness is constant [18]. Though this assumption simplifies somewhat the equation of motion at each node, it is undesirable for accurate simulation of the real tissue that has a nonlinear stress-strain relationship. Further more, there is a tendency



Fig. 4: An illustration of the movements of driven points in different facial expressions.

of mass-spring to be maximum or minimum under relatively large compression, which is due to the fact that linear springs used can be compressed fully. It is natural to investigate the problem of the elastic stiffness calculation for nonlinearity factor varying with muscle deformation. Thus we need to discuss the nonlinear mass-spring model in the following sub-section.

# A. Nonlinear Mass-spring Model for Facial Muscle Deformation

As discussed in [19], the mechanical law of soft-issue points is modeled by a nonlinear function. Thus we employed their approach to describe the deformation of the facial muscles and calculate the elastic stiffness and elastic force for each facial muscle. Suppose that  $x_i$  is an arbitrary driven point, and  $x_j$  denotes the corresponding fixed point of  $x_i$ . The length of a facial muscle in the neutral state is denoted by  $d_{ij}$ . Let  $\Delta x_{ij} = x_i - x_j$ , function  $K(x_i, x_j)$  is introduced to modulate the constant elastic stiffness  $k_0$ :

$$K(x_i, x_j) = (1 + (|\Delta x_{ij}| - d_{ij})^2)^{\alpha} k_0 \qquad (1)$$

and the elastic force generated by an spring is:

$$f(x_i, x_j) = K(x_i, x_j) \frac{(|\Delta x_{ij}| - d_{ij})}{|\Delta x_{ij}|} \Delta x_{ij}$$
(2)

In equation (1),  $\alpha$  is the nonlinearity factor controlling the modulation. In the later sections, for clarity let  $f_{(ij)} \triangleq f(x_i, x_j)$ . By assigning different values to  $\alpha$ , function  $K(x_i, x_j)$  can be chosen to model linear or nonlinear stress-strain relationship. For example,  $\alpha$  is taken the value of zero for the linear spring model or non-zero for the nonlinear spring model.

## B. Elastic Force

In this section, we study the characteristics of elastic forces of facial muscles for different facial expressions, and then extract the novel visual features based on such characteristics for facial expression recognition. The psychological experiments as shown in [20] have suggested that facial expressions are more accurately recognized from a temporal behaviors from a single static image. The temporal information often reveals the underlying emotional states. Therefore, our work concentrates on modeling the temporal behaviors of facial expressions from their dynamic appearances in an image sequence. Fig. 5 shows the image sequences on three kinds of facial expressions, namely happy, surprise and sadness. All the sequences start from the neutral state and transit to the emotional state.



Fig. 5: Video examples on facial expressions.

Figure 6 shows the variations of elastic forces of different facial muscles during these three facial expressions. The x-axis denotes the time, and the y-axis denotes the magnitudes of elastic forces. If the magnitude of elastic force is more than zero, the facial muscle is being contracted. If the magnitude of elastic force is less than zero, the facial muscle is being stretched. Due to the reason of symmetry, we only show the variations of elastic forces of 13 facial muscles, namely fl1, fl2, fl3, el, nl, cl, ml1, ml2, lp1, lp2, lp3, lp4 and jl.

In terms of the variations of elastic forces in Figure 6, there are always three phrases for an facial expression, namely starting phase, apex phase and ending phase. At the neutral state, all the driven points of facial muscles locate at their original position, and the elastic forces of facial muscles are all equal to zero. When one facial expression reaches its apex phase, the magnitudes of the elastic forces will also reach the largest values. When the expression is transiting to the ending state, the magnitudes of elastic forces will decrease accordingly. We also observed that most of the largest magnitudes of elastic forces are different for different facial expressions. Thus, we extract the largest magnitudes of elastic forces of facial muscles as the facial features for facial expression recognition.

# C. Facial Expression Recognition

In the system, facial expression recognition is formulated as a classification problem. The input for the classification module is a 22 dimension vector, and each element denotes the magnitude which has the largest absolute value during a facial expression. To classify the input vectors, we employ



Fig. 6: An illustration of the variations of elastic forces for three facial expression in Fig. 5

SVM as the classifier, since it is able to construct arbitrary decision boundaries[21].

# V. EXPERIMENTAL RESULTS ON FACIAL EXPRESSION RECOGNITION

In the system, the resolution of the acquired images is  $320 \times 240$  pixel. The system is developed under Microsoft by using Visual Studio . NET 2005. OpenCV [15] is employed to implement the module of face detection and key point extraction. To evaluate the system for facial expression recognition, we generate a total of 600 videos for six facial expressions (100 videos for each facial expression), namely happy, sad, fear, disgust, anger and surprise. In this paper, one video corresponds to one facial expression and consists of an image sequence. All the facial videos are automatically

captured from one person, since we do not touch the problem of face recognition. Then, all the data are divided into two groups randomly, 480 for training and 120 for testing. Thus we have 80 training data and 20 testing data for each facial expression class.

TABLE 1: Classification Results Using Nonlinear Mode

Emotion	Нарру	Sadness	Fear	Disgust	Anger	Surprise
Happiness	0.842	0	0.126	0.032	0	0
Sadness	0.009	0.733	0.153	0.070	0.035	0
Fear	0.054	0.063	0.706	0.023	0	0.154
Disgust	0	0.173	0.076	0.616	0.135	0
Anger	0	0	0.005	0.133	0.862	0
Surprise	0	0	0.088	0	0	0.912

$T \cap D \cap D \cap Z$ . Classification results Using Different with	TABLE 2:	Classification	<b>Results</b> Using	Linear	Mode
--	----------	----------------	----------------------	--------	------

				0		
Emotion	Нарру	Sadness	Fear	Disgust	Anger	Surprise
Happiness	0.632	0.083	0.219	0.052	0.004	0.010
Sadness	0.038	0.570	0.186	0.113	0.093	0
Fear	0.051	0.141	0.498	0.020	0.013	0.277
Disgust	0	0.014	0.132	0.561	0.287	0.006
Anger	0	0.011	0.036	0.251	0.702	0
Surprise	0.023	0.039	0.122	0.021	0.004	0.791

To evaluate the accuracy of facial expression recognition, all the results are tabulated in Table 1. We set  $\alpha = 1$  for the nonlinear model and  $\alpha = 0$  for the linear model. As shown in Table 1 and 2, the system based on nonlinear massspring model achieved better performance for all the facial expressions than the linear model. In particular, nonlinear model achieved the significant improvements for happy, sad, fear and surprise compared with linear model. This indicates two folds: 1) Nonlinear mass-spring model is more reasonable for describing the movements of facial muscles compared with the linear model. 2) Our proposed novel features based on elastic forces derived from nonlinear spring model are effective for facial expression recognition.

## VI. INTERACTIVE ROBOT EXPRESSION IMITATION

Facial expression recognition and imitation is an effective way for a social robot to understand human emotions and communicate with human beings, which plays a major role in human interaction and nonverbal communication. In order to build the effective communications between human and robots, an intuitive and natural method to implement an expressive robotic face which can imitate human emotions. Without loss of generality, we build an interactive robot expression imitation module based on our proposed approach for facial expression recognition. As shown in Fig. 7, a robot head is designed to imitate the human facial expressions. The input to the system is a video stream capturing the user's face.



Fig. 7: The robot head.

### A. Expressive Robotic Face

The robot head consists of 16 Degrees of Freedom (DOF) to imitate the facial expressions. The development of the expressive robotic face is further sub-divided into:

- The mechanical design of the robotic face, whereby the various components making up the robotic face are designed. Considerations are given to the joint and motor placement to produce different facial expressions.
- The software control of the servo motors. The motors are controlled through the New Micros ServoPod, which provides the PWM signals to the 16 servo motors. The program is developed using the New Micros operating system/language IsoMax .

A methodology for facial motion clone is developed, that is to copy a whole set of morph targets from a 2D real face image to an expressive robotic face. The inputs include two face images, one is in neutral position and the other is in a position containing some motion that to be animated, e.g. in a laugh expression. The target face model exists at the neutral state. The goal is to obtain the target face model with the expression copied from the source face. Based on the feature tracking method we described before, the tester's facial features vector at the neutral state is subtracted from that at the expression. Therefore, the displacement and velocity information are extracted. They are multiplied by the a weight vector to reach the desired animation effects, e.g. exaggerated expression. The weight vector can be predefined according to the desired animation effects. Subsequently, the weighted vector is added on the face plane of the robot head in its neutral state. The robot head is able to show its emotions through an array of features situated in the frontal part of the head.

### B. Generation of Artificial Facial Expression

The facial expression generation is based on Ekman's six basic emotions(happiness, surprise, sadness, disgust, fear, anger) [16], [22]. In the system, the robot can imitate six human facial expressions plus the neutral state with no expressions.

In terms of our design, the robot head are triggered to imitate human facial expressions by the emotion generator engine, and can generate a vivid imitation according to the tester's facial expression. For instance, our robot can imitate the happiness once it detects a facial expression of happiness (through the vision system). This application is a very simple one, in which the robot is just imitating the expression of a human subject. In other words, to see its reaction according to the emotional state displayed by a person. Usually, the response of the robot occurs slightly after the apex of the human expression. The results of this application were recorded in a 2 minute video. In order to be able to display simultaneously in the video the correspondence between person and robot expressions, we put them side by side. In this case only, we analyzed offline the content of the video and commands with the facial expression code were sent to the robot. Fig. 8 illustrates nine detected keyframes from the frame video. These are shown in correspondence with the robots response. The middle column shows the recognized expression. The right column shows a snapshot of the robot head when it interacts with the detected and recognized expression.



Fig. 8: Left column: Some detected keyframes associated with the video. Middle column: The recognized expression. Right column: The corresponding robot's response.

## VII. CONCLUSIONS

In this paper, we proposed an interactive system for recognizing and imitating human facial expressions. In the system, nonlinear mass-spring model was employed to simulate twenty two facial muscles' deformations during facial expressions, and then the elastic forces of the facial muscles' deformation were taken as the novel features to be grouped into a vector. Then such vectors were input into the module of facial expression recognition. The experimental results showed that our proposed nonlinear facial mass-spring model coupled with the SVM classifier is effective to recognize the facial expressions compared with the linear mass-spring model. At the back end of the system, a social robot was designed to make artificial facial expressions. Experimental results of facial expression generation demonstrated that our robot can imitate six types of facial expressions effectively. In practice, six facial expressions are not enough to reflect human emotions. For example, hot anger and cold are two different anger expressions. Thus we will define more facial expressions and improve our proposed system to accurately

#### recognize and imitate more facial expressions in the future.

### REFERENCES

- S. S. Ge, "Social Robotics: Integrating Advances in Engineering and Computer Science," *Proceedings of Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology International Conference*, pp. xvii–xxvi, May 9-12 2007. Chiang Rai, Thailand.
- [2] G. Y. Park, S. Lee, W. Y. Kwon, and J. B. Kim, "Neurocognitive affective system for an emotive robot," 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2595–2600, October 2006.
- [3] S. S. Ge, Y. Yang, T. H. Lee, and C. Wang, "Facial expression recognition and tracking based on distributed locally linear embedding and expression motion energy," *will appear in Journal of Intelligent Service Robotics -Special Issue*, 2008.
- [4] A. W. Young and H. D. Ellis(Editors), Handbook of Research on Face Processing. North-Holland, Amsterdam: Elsevier Science Publishers B.V., 1989.
- [5] C. Padgett and G. Cottrell, *Representing face images for classifying emotions*, vol. 9. Cambridge, MA: MIT Press, 1997.
  [6] Y. Yacoob and L. S. Davis, "Recognizing human facial expressions"
- [6] Y. Yacoob and L. S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, pp. 636–642, June 1996.
- [7] B. Fasela and J. Luettinb, "Automatic Facial Expression Analysis: A Survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.
- [8] B. Abboud, F. Davoine, and M. Dang, "Facial expression recognition and synthesis based on an appearance model," *Signal Processing: Image Communication*, vol. 19, no. 8, pp. 723–740, 2004.
  [9] P. Roivainen, H. Li, and R. Forcheimer, "3-D motion estimation in
- [9] P. Roivainen, H. Li, and R. Forcheimer, "3-D motion estimation in model-based facial image coding," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, pp. 545–555, 1993.
- [10] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying Facial Actions," *IEEE Transactions on Pattern Analysis* and Machine Intelligence, vol. 21, no. 10, pp. 974–989, 1999.
- [11] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 36, pp. 96–105, February 2006.
- [12] H. Miwa, K. Itoh, M. Matsumoto, M. Zecca, H. Takanobu, S. Roccella, M. C. Carrozza, P. Dario, and A. Takanishi, "Effective emotional expressions with emotion expression humanoid robot we-4rii," *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 2203–2208, Oct. 2004.
- [13] http://www.ai.mit.edu/projects/humanoid-robotics-
- group/kismet/kismet.html.
  [14] Y. Zhang, E. C. Prakash, and E. Sung, "Hierarchical Face Modeling and Fast 3D Facial Expression Synthesis," *Proceedings of the XV Brazilian Symposium on Computer Graphics and Image Processing* (SIBGRAPI02), pp. 357–364, October 7-10 2002. Fortaleza, Brazil.
- [15] Intel Corporation, *OpenCV Reference Manual*, 2001. http://www.intel.com/technology/computing/opencv/index.htm.
- [16] P. Ekman and W. V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, California, USA: Consulting Psychologists Press, 1978.
- [17] S. Platt and N. Badler, "Animating facial expressions," Proc. SIG-GRAPH81, ACM SIGGRAPH, vol. 15, pp. 245–252, 1981.
- [18] Y. Fung, Biomechanics: Mechanical Properties of Living Tissues. New York: Springer Verlag, 1993.
- [19] Y. Zhang, E. C. Pracash, and E. Sung, "A new physical model with multilayer architecture for facial expression animation using dynamic adaptive mesh," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, pp. 339–352, May/June 2004.
- [20] J. N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *Journal* of *Personality Social Psychology*, vol. 37, pp. 2049–2059, 1979.
  [21] I. Kotsia and I. Pitas, "Facial expression recognition in image se-
- [21] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, vol. 16, pp. 172– 187, January 2007.
- [22] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System*. Salt lake City, USA: A Human Face, 2002.