Face Recognition by Efficient Facial Expressions for Human-Machine Interaction

Neha Agrawal[1], Rishabh Agrawal[2], Sushil Kumar[1]
[1]Technocrats Institute of Technology (TIT), Bhopal, [2]Blue Star Ltd.

Abstract

In this paper approaches for facial expression analysis and face recognition. First, we propose an efficient facial expression recognition scheme based on the detection of key frames in videos where the recognition is performed using a temporal classifier. Second, we use the proposed method for extending the human-machine interaction functionality of the robot. More precisely, the robot is displaying an emotional state in response to the recognized using facial expressions. Recognition is performed via a 'Simple Recurrent Network' which lends itself well to modeling dynamic events in both users' face recognition through facial expressions. This approach works in models using expressivity's by a dimensional representation of detecting the usual 'universal emotions' which are happiness, sadness, fear, disgust, anger, surprise works in everyday human-machine interaction. Facial expressions is a 'personalized' facial expression recognition system, works on facial expression recognition focus ***on*** *the Well-known Six universal facial expressions(happy, sad, fear, angry, surprise and disgust) classification approach.*

Keywords: System description, personalized, facial mass spring model, feature extraction and facial expression recognition, expressive robotic face, fuzzy neural network.

1. 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]. The algorithm is deployed on an database which was recorded simulating human-machine discourse and, therefore, contains less extreme expressivity's and subtle variations of a number of emotion labels. 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. Facial expression recognition is very important in many

Human -robot human- computer interaction **System.** Finally, we introduce our robot that can make artificial 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 expressions 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 human facial expressions [2].

Within the recent decade, many researchers have been trying to automatically recognize facial expressions of human beings. Various pattern recognition methods have been used to recognize facial expressions. Facial expression recognition approaches could be divided into two main categories: target oriented and gesture oriented. Target oriented approaches attempt to infer the facial expression only from one typical facial image [3], [4].

On Facial Action Coding System (FACS) model in [5]. 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 emulated by a set of mass-springs, i.e. each edge of the mesh can be modeled by a linear spring. Another area of research on human robot interaction (HRI) is for robot to generate facial expressions. Currently, one of representative HRI systems is the mechanical looking robot.

The existing 3D face mesh for facial expression recognition is based on the assumption of linear mass-spring model. As discussed in [6], the simple linear mass-spring models cannot emulate the real tissue/muscles accurately. Thus we employ nonlinear mass-spring model to describe 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.

2. 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.

'Feature Extraction' stage is passed to the second 'FNN-based classifier' stage **as** an input.

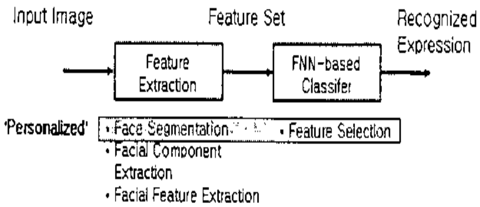

Fig1.Strucrure of 'personalized' facial expression recognition system

A. Personalized

'Personalized' concept can be dealt with personalized classification approach for each person's expressions. In view of classifier design, we define the term 'personalization' as follows. From a given general structure of classifier; according to the characteristics of given input/output pairs, a specifically Structured classifier can be generated by 'personalization' process. Besides, each generated classifier should be composed of different starchier" one to one. Here, 'personalization' is implemented by different input features for classifier. That is, for individual person, a set of input features for facial expressions can be varied according to his/her facial structure, race, culture, age, sex and so on. Among many pattern recognition-related works, feature selection is one of the possible solutions for this purpose. The effectiveness of feature selection process in the complex pattern recognition problems [7][8][9][10][11]. Thus, in this paper, we propose a novel feature selection method for 'personalized' facial expression recognition. As a fundamental classifier, we also use fuzzy neural networks (F")-Based classifier among many soft computing techniques.

B. Facial mass spring model

For Facial Action Coding System (FACS) in [12], 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. The module of the system 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 is assumed to consist of a pair of key points, namely driven point, fixed point and the facial mass-spring model. The fixed points cannot be moved during a facial expression.

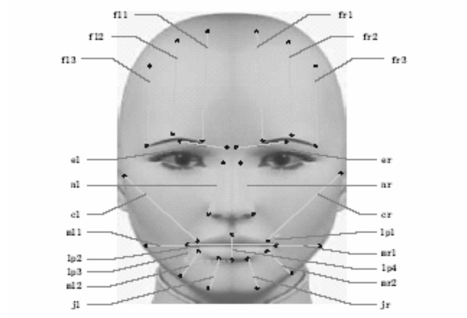

Fig2. The facial mass spring model

Given the outputs of feature extraction and a predefined set of facial expressions, the classification module classifies 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. Include 'personalized' concept for face segmentation process and feature selection process in each step.

3. 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 emulation of the real tissue that has a nonlinear stress-strain relationship. Further more, there is a tendency 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 non linearity factor varying with muscle deformation

A. Nonlinear Mass-spring Model for Facial Muscle Deformation

As discussed in [19], the mechanical law of soft tissue 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 a constant elastic stiffness k0:

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

and the elastic force generated by spring is:

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

In equation (1), α is the no linearity factor controlling the modulation. In the later sections, for clarity let f (ij) $\triangleq$ f (x_i, x_j). By assigning different values to α, function K (x_i, x_j) can be chosen to model linear or nonlinear stress strain relationship. For example, α 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

C. Facial Feature Extraction

According to previous step, the locations of each facial component are found. These facial components are categorized as 'permanent' facial components which are constantly appeared none the less the change of facial expressions. On the contrary, the 'transient' components are appeared or disappeared according to the change of facial expressions. Fig. 3 shows a typical set of transient components [12].
From above two types of facial components, meaningful features should be extracted for facial expression recognition

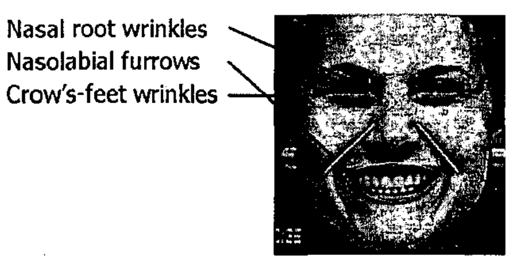

Fig3. Transient Facial components

Among many possible features, we select five features according to their importance for facial expressions based on previous related work **[17].** In this paper, the selected features are
A: degree of mouth openness, B: degree of eye' openness, C: the vertical distance between the eyebrows and the eyes, D: degree of nasolabial root wrinkles (NLR) and E: degree of nasolabial furrows (NLF). For C, D and E, which are transient components, we use well-known image features such as Euclidean distance(C) and Gabor-filtered coefficients and E) [13**].**
To acquire effective image features for **A** and **B,** we use a human visual system-based approach. Human can extract meaningful features for eyes and mouth based on the combination of global and local features without much effort [14].
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 Servo Pod, which provides the PWM signals to the 16 servo motors. Therefore we can use IsoMax, which is the New Micros operating system language, to implement the action units [15] or imitate the facial expressions by controlling those servo motors. For example, Mouth stretch can be imitated by controlling two servo motors of upper lip and lower lip.

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 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.

Degree of mouth openness:

The degree of mouth openness can be measured by combining global features (the height ratio and the area ratio between the whole face and the mouth

region) and a local feature. For example, if one person has
very thick lips than others, we cannot easily know the degree of .mouth openness just using global features. **As** a local feature, we use the Gabor-gaussian feature (see (1)) [16]. Intuitively, this feature represents the degree of mouth openness by considering vertical displacement of lips and vertical edge information. This type of feature is robust to varying illumination, change of lip shapes and ***so*** on.

$$f_G = \frac{\sum_{j=1}^{H-1} w_G(j) dy_{proj}(j)}{\sum_{j=1}^{H-1} w_G(j)},$$

Where, w_G**(j)** means the gaussian weights, $\mathbf{dy_{proj}(j)}$ is the absolute values of the derivative of the projected value $\mathbf{y_{proj}(j)}$, **H** is the height of gator-filtered image.

Degree of eye openness:

Next, for the degree of eye openness, we investigated various eye images and found that the shape of the upper eyelid is clearly different between closed eyes and opened eyes. Thus, by acquiring the shape information of the upper eyelid, we can easily estimate the degree of eye openness. For this purpose, we use the 'dip' feature **[17]** in Log-polar mapped image. As shown in Fig. 4, 'dip' features defined as the vertical distance between the base line to the target segment (in this case, the target segment is shown as 'Upper eyelid' in Fig. 4). Similarly in case of mouth openness, the dip feature acts **as** a global feature. For a local feature, we also use Gabor-filtered coefficients between upper eyelid and lower eyelid.

Fig4. Feature extraction for eye openness

4. Facial Expression Recognition Using Fuzzy Neural Networks (Fnn)

FNN is a kind of neural networks-based implementation of fuzzy decision making system. It is characterized by the advantages of both systems such as knowledge representation of human experts and learning function of neural networks [18].FNN mainly consists of 5-layered structure such as; 1) 'Input Variable Node(or Input Feature Node)', ***2)*** 'Input Linguistic *Node',* 3)'Rule Node', 4) 'Output Linguistic Node', 5) 'Output Variable Node'. Input/Output Variable Nodes are similar concept as the conventional neural networks. The second layer is for representing input linguistic terms for each input variable node, and the fourth layer is for representing output linguistic terms for each output variable node. Each input/output linguistic nodes plays a function to produce the membership value between 0 and 1 with well-known bell-shaped gaussian function defined by parameters stands for center location and standard deviation in the universe of discourse. In the third layer, each node connects the nodes in the second layer with the nodes in the fourth layer. This structural relationship corresponds to fuzzy logic-based inference with antecedents (the nodes in the second layer) and consequents (the nodes in the forth layer). In Fig. 5, a simple FNN is given with notations. Here, two inputs are 'X1' and 'X2' and two outputs are 'y1' and 'y2'. Two
Linguistic terms u e defined for each input/output node with subscripts 'L'and 'H'.

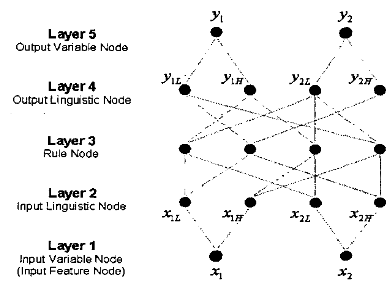

Fig5.Simple example for FNN 5-layered structure

In the facial expression recognition, Ekman's work [19] is considered as a starting point to construct FNN-based classifier for individual. By adding input/output patterns for each person's different facial expression,

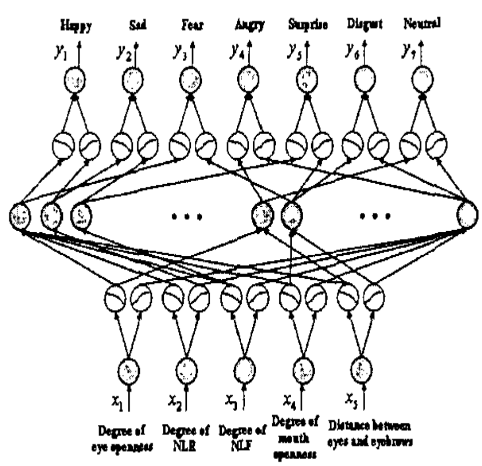

Fig 6 .Normal facial expressions using FNN

5. Recognition Results

With predefined image features and output classes, we construct FNN-based classifier as shown in Fig. 5. Here, output Classes are defined as six universal facial expressions [19] plus one normal facial expression (see Fig. 6). As a reference of facial expression, we use well-known persons [19]. Based on this reference DB, as training/test data set, we define several data sets. One is the same as Ekman's data sets are artificially generated from the original Ekman's work. These data sets are generated from the input image features of FNN with various standard deviations. Specifically, we use standard deviation values such as 1%, 2%, ***5%,*** 10%. 20% and ***40%*** respectively. Thus, six data sets are generated and each data set contains input/output patterns (100 input patterns are generated for a person, respectively). To evaluate our proposed method, we use several numerical measures such as the size of the network (the number of nodes and branches), the execution time and the recognition rate and comparison with for various test sets give accurate and precise results to get human-machine interaction.

6. Conclusion

In this paper, an interactive system has been developed for recognizing face and human facial expressions for human-machine interface (Fig.7)

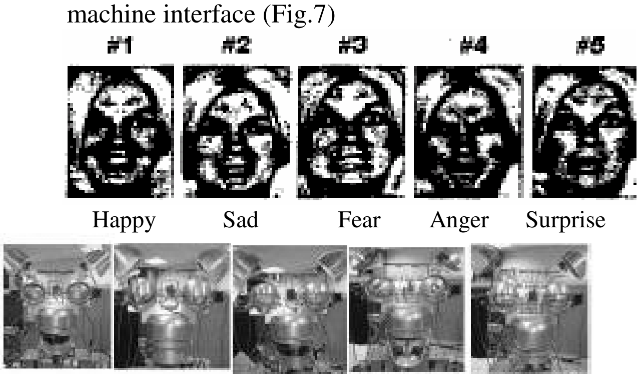

Fig7. Human Machine Interaction through Facial Expression

In the system, nonlinear mass-spring model is 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 are input into the module of facial expression recognition. The experimental results showed that our proposed nonlinear facial mass-spring possible solution is to investigate user's responses to the model coupled with the SVM classifier is effective with FNN approach 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 recognize face by six types of facial expressions effectively. Until now, there is no results to explain how to estimate the model parameter α and k0, investigate is still a problem in our future work. We currently could not evaluate the expression quality of the proposed robot head, so one facial expressions of the proposed robot.

7. Acknowledgment

Neha agrawal thanks to Mr. Sushil Kumar for his valuable guidance and also credit to Mr. Rishabh Agrawal for his technical help and support in the corresponding paper.

8. References

[1] 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] 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.

[3] A. W .Young and H. D. Ellis(Editors), *Handbook of Research on Face Processing*. North-Holland, Amsterdam: Elsevier Science Publishers B.V., 1989.

[4] C. Padgett and G. Cottrell, *Representing face images for classifying emotions*, vol. 9. Cambridge, MA: MIT Press, 1997.

[5] 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, July

[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]K. Chug . "Performance Comparison of S y r a l Feature Selection Methods Bared on Node Pruning in Handwritten Character Recognition," *Proceedings* ofICDAR'99, pp. 11-15, 1997.

[8] V. Omia , "Feature Selection method wing neural networks:' *Proceedings ofICIPZW1*.pp. 513-516,2001.

[91 D. Chahrabarty , "Designing de-based classification with on-line feature selection: a neuro-fuzzy approach:' *Proceedings of AFSS2WZ.* pp. 251-259.2002.

[I0]:s. Han. W.-C. Bang and Z. Zenn Bien, "New Feature Set Selection Algorithm based on Soft Computing Techniques and its Application to EMF Pattern Classifcation:' Joumal of Fuzzy Optimization and Decision Making, Vol. 8. No. 3. pp. l-30,2002.

[11]M.-1.Kim.J.-S.Han.K.-H.Park,W.-C. BangandZ.ZeMBien," Classifcation of Arrhythmia Based on Discrete Wavelet

Transform and Rough Set Theory," ICCAS 2001. October 17-21,ZWI.ne 1996.

[12] 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.

[13] 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.

[14] 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.

[15] Y. M. Zhang and Q. Ji, "Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 699–714, 2005.

[I6] Z. Zhang Comparison between geometry-based and gabor wavelets-based facial expression recognition using multi-layer perception", IEEE International Conference *on* Automatic *Face and* Gesture Recognition. pp. 454459. 1998.

[17] G. Donato "Classifying facial actions". *IEEE Transactions on PAMI,*pp. 974-989, Vol. 21. No. IO. 1999. [18] [201 D.-1. Kim "Effective intention reading technique as a means ofhuman-robot interaction for human centered system", *IEEE International Conference on Fuzzy-* System, S303, 2001.

[19] P. Ekman al. "Facial Action Coding System". 1978.

[20] A. J. Fridlund, Human Facial Expression: An Evolutionary Perspective, Academic Press, 1994.