

Spontaneous Human-Robot Emotional Interaction Through Facial Expressions

Ali Meghdari¹(✉), Minoo Alemi^{1,2}, Ali Ghorbandaei Pour¹,
and Alireza Taheri¹

¹ Social and Cognitive Robotics Laboratory, Center of Excellence in Design,
Robotics and Automation (CEDRA), Sharif University of Technology,
Tehran, Iran

meghdari@sharif.ir, alemi@sharif.edu

² Islamic Azad University, Tehran-West Branch, Tehran, Iran

Abstract. One of the main issues in the field of social and cognitive robotics is the robot's ability to recognize emotional states and emotional interaction between robots and humans. Through effective emotional interaction, robots will be able to perform many tasks in human society. In this research, we have developed a robotic platform and a vision system to recognize the emotional state of the user through its facial expressions, which leads to a more realistic human-robot interaction (HRI). First, a number of features are extracted according to points detected by a vision system from the face of the user. Then, the emotional state of the user is analyzed with the help of these features. For the decision making unit, a state machine is designed that utilizes the results obtained from the emotional state analysis to generate the robot's response. Finally, a fuzzy algorithm is used to improve the quality of emotional interaction and the results are implemented on a commercial humanoid robot platform which has the ability of producing facial expressions.

Keywords: Social robotics · Human-robot interaction (HRI) · Emotional state recognition · Fuzzy finite state machine · Fuzzy clustering

1 Introduction

Since the creation of the first robot, researchers have been interested in development of interaction between a robot and its environment, with the possibility of robots interacting with each other and with humans. The common assumption is that humans prefer to interact with machines in the same way that they interact with other people. In this regard, different ideas and prototypes of robotic heads have been developed for HRI purposes [1–4]. Tadesse et al. [1] designed and implemented a twelve degrees-of-freedom humanoid baby head, capable of producing 6 basic facial expressions, and Saffari and Meghdari et al. [2] introduced a robotic head which turns toward the speaker in noisy environments. By improving the abilities of humanoid robots, they now have the capability to enhance scenarios involving education [5, 6], physical therapy [7–9], and elderly care [10]. In this regard, social learning and imitation, gesture and natural language communication, emotion, and recognition of interaction

partners are all important factors. In recent years, this field has attracted considerable attention from academic and the research communities. Zacharatos et al. [11] described recent emerging techniques and advances in automatic emotion recognition. Halder et al. [12] used an interval and a general type-2 fuzzy set separately to model the fuzzy face space for emotion recognition purpose.

In general, one can classify HRI studies into verbal and non-verbal interactive communications [13]. Figure 1 shows a model of an emotion-based HRI system. Aly et al. [14] introduced a multimodal behavior HRI for more naturally emotional interaction. A group of studies has been done based on emotional state detection through voice analysis [15]. There are also remarkable studies on emotion recognition according to the user's gestures [16–21]. Xiao et al. [16] involved a set of 12 upper body gestures to communicate with the robot. Chakraborty et al. [17] proposed a simple and robust scheme for emotion recognition and control, with good accuracy based on fuzzy relational approach, and geometric deformation facial features has been also used for facial expression recognition [18].

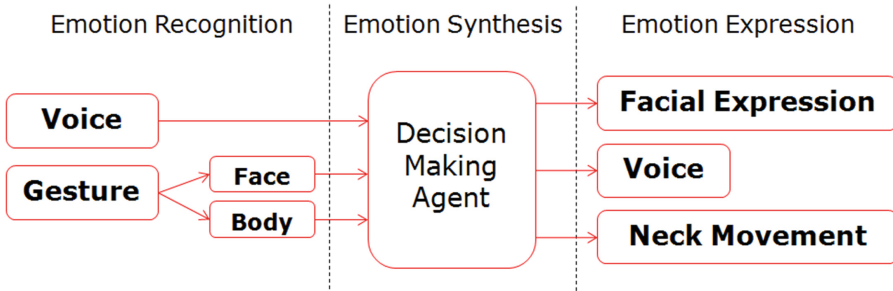


Fig. 1. Model of emotion-based HRI for humanoid robotic platforms

This paper presents an initial attempt to develop a robotic platform for social interaction research. This platform has an attractive physical appearance with which humans should enjoy interacting. Our emotion recognition is based on the user's facial gestures, and the robot's response is through facial expression and neck movement, in accordance to a fuzzy decision making algorithm. The desired work is the synchronization between the developed interaction mode and the implementation of the proposed emotion-based control.

2 Instruments

2.1 A Humanoid Robot

R50 – Alice, with the Iranian name “Mina”, is a humanoid robot made by Hanson Robokind Company, designed specifically for human-robot social interaction and has been used widely for studies on developmental and social robotics [22]. Mina is 69 cm high, weights 5.7 kg and has 32 degrees-of-freedom. Mina has the 3D face of a girl,



Fig. 2. The R-50 Alice (Mina) robot

which permits 11 degrees-of-freedom (Fig. 2) for generating facial expressions such as surprise, anger, happiness, sadness and so on. She also has 3 degrees-of-freedom in her neck which makes her able to trace the user by moving her head toward the user's face while they are in the interacting mode.

2.2 Machine Vision

In this study, we have used a Microsoft Kinect Sensor for our Machine vision application. The Microsoft Kinect Sensor is a physical device with depth sensing technology, a built-in color camera, and an infrared emitter that can sense the location and movement of people. With the help of version 2 from the Kinect for Windows Software Development Kit (SDK) it is possible to access a list of face points to extract our features. The positions of these points are defined in the Kinect body coordinate system. The origin is located at the optical center of the camera, the Z-axis is pointing towards a user, the Y-axis is pointing up, and the X-axis is to the right [23].

3 Research Methodology

3.1 Face Feature Extraction

In the first step, we have chosen 21 face points among the available 36 points detected in SDK (Table 1). These points were chosen in such a way to define facial features based on the action units of the Facial Action Coding System (FACS) [24]. Afterwards, a set of 18 features were defined according to changes in the distances between these points. These features are listed in Table 2. The data recorded from Kinect output and each feature is updated with a speed of 30 frames per second. In order to reduce the effect of noise on the extracted features, a moving average filter with a period of 5 previous data points is applied to each feature. Another issue is that the subject's features should be scale invariant, for this reason and to avoid the effect of user's distance from the Kinect Sensor (normalizing the features) all of these features are divided by the length of the subject's nose (our 18th feature). After making our features scale invariant, the first 17 normalized features are used to detect the emotional state of the user.

Table 1. List of the face points used for feature extraction

Face points	Description
P1	Inner corner of the left eye
P2	Outer corner of the left eye
P3	Inner corner of the right eye
P4	Outer corner of the right eye
P5	Inner left eyebrow
P6	Outer left eyebrow
P7	Center of the left eyebrow
P8	Inner right eyebrow
P9	Outer right eyebrow
P10	Center of the right eyebrow
P11	Left corner of the mouth
P12	Right corner of the mouth
P13	Middle of the upper lip
P14	Middle of the lower lip
P15	Bottom of the nose
P16	Bottom left of the nose
P17	Bottom right of the nose
P18	Left cheek bone
P19	Right cheek bone
P20	Left end of the lower jaw
P21	Right end of the lower jaw

Table 2. List of the facial features

Face features	Description
F1	Distance between P1 and P5
F2	Distance between P2 and P6
F3	Distance between P7 and the line crossing P1 and P2
F4	Distance between P3 and P8
F5	Distance between P4 and P9
F6	Distance between P10 and the line crossing P3 and P4
F7	Distance between P5 and P8
F8	Distance between P11 and P12
F9	Distance between P13 and P14
F10	Distance between P11 and P1
F11	Distance between P11 and P18
F12	Distance between P11 and P20
F13	Distance between P12 and P3
F14	Distance between P12 and P19
F15	Distance between P12 and P21
F16	Distance between P1 and P16
F17	Distance between P3 and P17
F18	Distance between P15 and the line crossing from P1 and P3

3.2 Emotional State Recognition

A data base of facial features of 3,000 samples was gathered from different poses for 6 main facial expressions from 10 different young adults (500 samples for each facial emotional state). These main emotional states are happiness, sadness, anger, surprise, disgust, and fear. Figure 3 shows some of our data base samples. This data base was used to train a fuzzy classifier which indicates the basic emotional state of the user through facial expression. For this reason, a fuzzy clustering method, called Fuzzy C-Means method (FCM) was used [25].



Fig. 3. Ten subjects from our database, posing facial expressions

In order to have more realistic samples, they are tried to be spontaneous facial expressions [20]. To reach this goal the sample was selected from a range of videos captured from our subjects, while they are expressing their emotion.

3.3 Producing Emotional Reaction

The first step toward a more realistic response is tracking the user by moving the head of the robot. For this purpose, Neck Yaw and Neck Pitch (angles for rotating in the azimuth and elevation planes) were adjusted such that Mina was always facing her user as she responded to the user's emotional state.

A smooth path was designed for each of these angles of turning. These angles were calculated according to the position of the user's head. From the data output of Kinect, the head position is available in the Kinect body coordinate system. This position needs to be transferred to the robot's head coordinate system, to calculate the proper angles. Figure 4 shows the head position in both coordinate systems and proper rotating angles in the robot's head coordinate system.

In the next step, a finite state machine was used to generate an emotional reaction, consisting of one of the six emotional states, to indicate the emotional state of the robot. The input to the state machine was the user's emotional state (Fig. 5). The output of

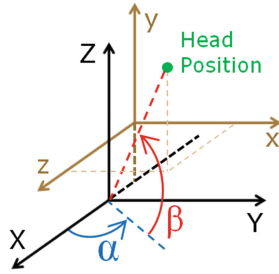


Fig. 4. Neck Yaw (α), Neck Pitch (β), Kinect body coordinate system (xyz), and Robot’s head coordinate system (XYZ)

each state was a set of facial expressions produced by Mina, declaring her reaction to the user’s emotional state. This output is set as a vector, containing the actuation level for each degree of freedom in the robot’s face. Since the robot is not supposed to become angry, there are no states considered for anger. Transition between states is according to the user’s detected emotional state.

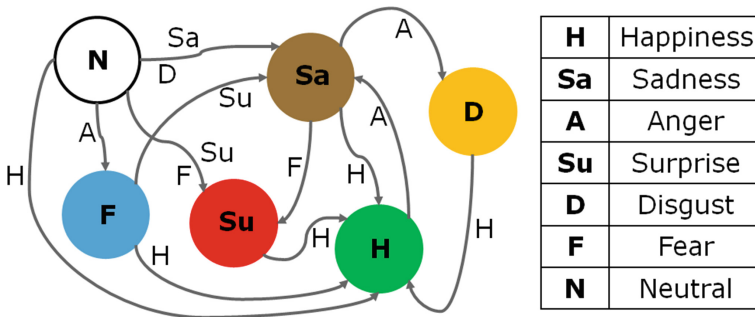


Fig. 5. The state machine diagram.

Since the algorithm used for emotional state recognition has a fuzzy output, a more realistic reaction can be generated by realizing the membership values of the user’s facial expression for each emotional state. Also, the state machine can be implemented with a number of if-then rules as follows:

```

if (state == 'N' && entry == 'H') then {state = 'H'}
if (state == 'N' && entry == 'A') then {state = 'F'}
if (state == 'N' && entry == 'D') then {state = 'Sa'}
if (state == 'N' && entry == 'F') then {state = 'Su'}
if (state == 'H' && entry == 'A') then {state = 'Sa'}
if (state == 'F' && entry == 'Su') then {state = 'Sa'}
if (state == 'Sa' && entry == 'F') then {state = 'Su'}
if (state == 'Su' && entry == 'H') then {state = 'H'}
...
    
```

These rules are taken as the rule base of our fuzzy inference system. A fuzzy inference system is a method that interprets the membership values in the input vectors and based on our defined rules, assigns values to the output vector [26]. Then, for the system entry and each state, a membership value is considered (in the beginning all of the state membership values are zero). By assigning the minimum of the membership value of the system entry and current state to the next coming state, and weighted averaging between the outputs of the states, a new level of emotional reaction is generated. For calculating the weighted average, states with the membership functions of more than 0.5 are taken into consideration.

4 Results

4.1 Feature Extraction and Emotion Recognition

Figures 6 and 7 show some face features evolution (after normalizing and noise filtering) during facial expression (the X axis is frame number and the Y axis illustrates change in face features). Features in all of these video sequences begin in a neutral state. As it can be seen, face features are defined in a way to be noticeably different for each facial expression, which leads to an easier and more accurate classification.

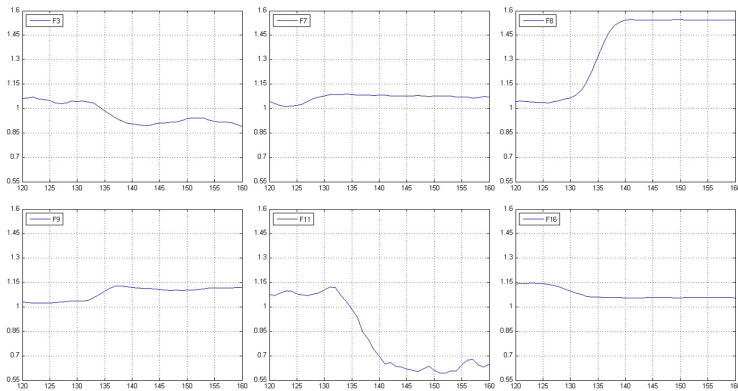


Fig. 6. Facial features variation from video sequence for happiness

In order to validate the classification process, another set of data, containing 700 samples from a new group of people (100 samples for each emotional state and 100 samples of neutral face) is used. The highest membership value indicates the emotional state of each sample. A sample is considered neutral if all of the corresponding membership values are less than 0.5. Table 3 presents the results from the test data. Each row indicates the detection results for each set of samples, with the same emotional state.

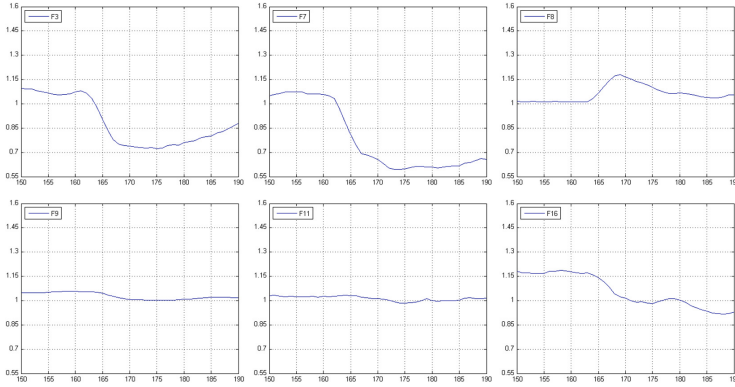


Fig. 7. Facial features variation from video sequence for anger

Table 3. Emotion recognition rate for test data

	Happiness	Sadness	Anger	Surprise	Disgust	Fear	Neutral
Happiness	98 %	0 %	0 %	1 %	0 %	0 %	1 %
Sadness	0 %	91 %	4 %	0 %	0 %	3 %	2 %
Anger	0 %	2 %	94 %	0 %	4 %	0 %	0 %
Surprise	0 %	0 %	0 %	99 %	0 %	1 %	0 %
Disgust	0 %	0 %	6 %	0 %	94 %	0 %	0 %
Fear	0 %	6 %	0 %	8 %	0 %	81 %	5 %
Neutral	0 %	3 %	0 %	0 %	0 %	2 %	95 %

4.2 Neck Movement

During the interaction Mina’s head turns to face the user. If the position of the user’s head moves while neck angles are moving toward their previous goal position, a new path will be generated according to the current neck angles values and the new destination angles (Fig. 9). Also, the new trajectory is considered to have the same

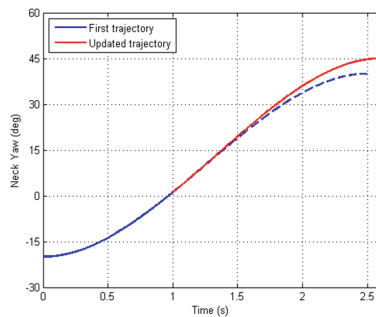


Fig. 8. Neck Yaw trajectory change in reaction to the change of user’s head position

velocity as the previous trajectory at the time the neck angles path changes its trajectory. This helps to have a smooth transition between trajectories.

4.3 Mina's Emotional Reaction

Using the fuzzy finite state machine for generating proper facial expressions caused more interacting modes, and a variable output level. Also, the change rate of the emotional state of the robot is dependent on the intensity of the user's facial expression. Figure 8 shows some of Mina's reactions to her user's emotional state.



Fig. 9. Generating combinatorial facial expressions by Mina

Since we wanted to develop this social robotic platform for further HRI applications, it was important to know the reaction of people and their impression about interacting with Mina. Therefore we attended two exhibitions with her. The feedback from the people interested to continue interacting with her was quite positive. Next, we are going to involve her in some intervention scenarios for autistic children as our future work.

5 Conclusions

Usually, emotional state is a combination of two or more basic emotions. To have a better HRI, detecting the share of each basic emotion in the users current emotional state is considered valuable. In this research, we detected the user's emotional state from his/her face gesture with fuzzy classification of extracted facial features. This method made it possible to assign a membership value to the facial expression of the user, meaning that the user's emotional state could be related to more than one basic emotional state. In addition, basic emotions were recognized as well with an overall accuracy of more than 90 % for 5 out of 6 basic emotions. Then, the identified facial expression was given to the state machine developed for emotional interaction. To expose the proper facial expression, Mina was programmed to turn her head to face the user. Finally, the HRI system was shown to be capable of producing a combinatorial facial expression output. The system was also able to decide and generate different facial expressions with variable intensities. As a result, Mina could communicate with human user more naturally.

References

1. Tadesse, Y., Hong, D., Priya, S.: Twelve degree of freedom baby humanoid head using shape memory alloy actuators. *J. Mech. Robot.* **3**, 211–226 (2011)
2. Saffari, E., Meghdari, A., Vazirnezhad, B., Alemi, M.: Ava (A Social Robot): design and performance of a robotic hearing apparatus. In: Tapus, A., André, E., Martin, J.-C., Ferland, F., Ammi, M. (eds.) *ICSR 2015. LNCS*, vol. 9388, pp. 440–450. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-25554-5_44](https://doi.org/10.1007/978-3-319-25554-5_44)
3. Hanumara, N.C., Slocum, A.H., Mitamura, T.: Design of a spherically actuated human interaction robot head. *J. Mech. Design* **134**(5), 055001 (2012)
4. Asfour, T., Welke, K., Azad, P., Ude, A., Dillmann, R.: The Karlsruhe humanoid head. In: 8th IEEE-RAS International Conference on Humanoid Robots, pp. 447–453, December 2008
5. Meghdari, A., Alemi, M., Ghazisaedy, M., Taheri, A.R., Karimian, A., Zandvakili, M.: Applying robots as teaching assistant in EFL classes at Iranian Middle-Schools. In: *Proceeding International Conference on Education & Modern Educational Technologies (EMET-2013)*, 28–30 September 2013, Venice, Italy (2013)
6. Alemi, M., Meghdari, A., Ghazisaedy, M.: The impact of social robotics on L2 Learners' anxiety and attitude in English vocabulary acquisition. *Int. J. Soc. Robot.* **7**(4), 523–535 (2015)
7. Alemi, M., Ghanbarzadeh, A., Meghdari, A., Moghaddam, L.J.: Clinical application of a humanoid robot in pediatric cancer interventions. *Int. J. Soc. Robot.* (2015)
8. Trinh, T.Q., Schroeter, C., Kessler, J., Gross, H.-M.: “Go Ahead, Please”: recognition and resolution of conflict situations in narrow passages for polite mobile robot navigation. In: Tapus, A., André, E., Martin, J.-C., Ferland, F., Ammi, M. (eds.) *ICSR 2015. LNCS*, vol. 9388, pp. 643–653. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-25554-5_64](https://doi.org/10.1007/978-3-319-25554-5_64)
9. Taheri, A.R., Alemi, M., Meghdari, A., PourEtemad, H.R., Basiri, N.M.: Social robots as assistants for autism therapy in Iran: research in progress. In: *Second RSI/ISM International Conference Robotics and Mechatronics (ICRoM)*, pp. 760–766. IEEE (2014)
10. McColl, D., Nejat, G.: A socially assistive robot that can monitor affect of the elderly during mealtime assistance. *J. Med. Dev.* **8**(3), 030941 (2014)
11. Zacharatos, H., Gatzoulis, C., Chrysanthou, Y.L.: Automatic emotion recognition based on body movement analysis: a survey. *IEEE Comput. Graph. Appl.* **34**(6), 35–45 (2014)
12. Halder, A., Konar, A., Mandal, R., Chakraborty, A., Bhowmik, P., Pal, N.R., Nagar, A.K.: General and interval type-2 fuzzy face-space approach to emotion recognition. *IEEE Trans. Syst. Man Cybern. Syst.* **43**(3), 587–605 (2013)
13. Mavridis, N.: A review of verbal and non-verbal human-robot interactive communication. *Robot. Auton. Syst.* **63**, 22–35 (2014)
14. Aly, A., Tapus, A.: Multimodal adapted robot behavior synthesis within a narrative human-robot interaction. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2986–2993 (2015)
15. Yashaswi Alva, M., Nachamai, M., Paulose, J.: A comprehensive survey on features and methods for speech emotion detection. In: *IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (2015)
16. Xiao, Y., Zhang, Z., Beck, A., Yuan, J., Thalmann, D.: Human–robot interaction by understanding upper body gestures. *Presence* **23**(2), 133–154 (2014)
17. Chakraborty, A., Konar, A., Chakraborty, U.K., Chatterjee, A.: Emotion recognition from facial expressions and its control using fuzzy logic. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **39**(4), 726–743 (2009)

18. Kotsia, I., Pitas, I.: Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Trans. Image Process.* **16**(1), 172–187 (2007)
19. Dahmane, M., Meunier, J.: Prototype-based modeling for facial expression analysis. *IEEE Trans. Multimedia* **16**(6), 1574–1584 (2014)
20. Li, Y., Mavadati, S.M., Mahoor, M.H., Zhao, Y., Ji, Q.: Measuring the intensity of spontaneous facial action units with dynamic Bayesian network. *Pattern Recogn.* **48**(11), 3417–3427 (2015)
21. Li, Y., Wang, S., Zhao, Y., Ji, Q.: Simultaneous facial feature tracking and facial expression recognition. *IEEE Trans. Image Process.* **22**(7), 2559–2573 (2013)
22. <http://www.robkindrobots.com/>
23. Kinect for Windows SDK (2016). <https://msdn.microsoft.com/en-us/library/>
24. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting psychologists press, Palo alto (1978)
25. Popescu, M., Keller, J., Bezdek, J.C., Zare, A.: Random projections fuzzy c-means (RPFCM) for big data clustering. In: *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, August 2015
26. Yan, J., Ryan, M., Power, J.: *Using Fuzzy Logic: Towards Intelligent Systems*, vol. 1. Prentice Hall, London (1994)