Proceedings of the 4th
International Conference on Robotics and Mechatronics
October 26-28, 2016, Tehran, Iran

The Real-Time Facial Imitation by a Social Humanoid Robot

Ali Meghdari*, Saeed Bagheri Shouraki, Alireza Siamy, Azadeh Shariati
Social & Cognitive Robotics Laboratory
Center of Excellence in Design, Robotics, and Automation (CEDRA)
Sharif University of Technology, Tehran, IRAN. *meghdari@sharif.edu

Abstract—**Facial expression imitation with applications in the design of human robot interaction (HRI) systems is an active area of research. In this study, we propose an approach for real-time imitation of human facial expression by a humanoid social robot "Alice". Artificial neural network (ANN) and Kinect sensor are used for recognition and classifying of the facial expressions like happiness, sadness, fear, anger and surprise; with the Alice humanoid robot imitating the comprehended expressions. Results and experiments demonstrate the effectiveness of the approach.**

Keywords*— *Artificial neural network (ANN), facial expression recognition, human robot interaction (HRI), imitation, Kinect senor, real time

I. INTRODUCTION

Human Robot Interaction (HRI) is a standout amongst the most critical undertakings in social robotics. Most of the HRI strategies which use non-verbal techniques are based on facial expressions [1-3]. They are similar to the way individuals interact in their everyday life. A man's face passes on vital data regarding feeling, mood such as happiness, sadness, fear, anger and surprise and numerous different signals related to anxiety or prosperity. Face is the record of brain and actions speak louder than words [4-6]. One of the active regions of research in HRI is to create robots with similar facial characteristics and representations to that of human. In other words, to create a robot that is capable of recognizing the facial expression and mimicking the human behavior and facial modes [7, 8].

Imitation of motions and emotions plays an essential role in the social robotic advancement [9, 10]. This paper tries to address the problem of facial expression recognition and real time mimicking imitation by considering: Kinect sensor for instantly recognize human face out of faces and non-faces objects; recognized the facial expressions of the detected face by using classification algorithm to be implemented on an Alice social humanoid; providing the robot with the ability to mimic the changes in the facial expressions, we aim to decrease the robot reaction time in order to enhance its interaction with human user. Throughout the years several facial recognition methodologies were proposed utilizing different elements and recognition techniques.

Song et al. utilized deep convolutional neural network to build up a face expression recognition system for smart-phone; they developed a real-time robust facial expression recognition function on a smartphone. To this end, they trained a deep convolutional neural network to classify facial expressions. The network has 65k neurons and consists of 5 layers [11]. Ebrahimi et al. used deep convolutional neural network (CNN) to analyze facial expression to perceive feelings in videos [12].

Barros et al. proposed a deep neural network model which is able to recognize spontaneous emotional expressions and to classify them as positive or negative [13]. Ijjina proposed facial expression recognition using deep CNN based on features generated from depth information [4]. Cid et al. exhibited a real-time framework for recognition and imitation of facial expressions with regards to affective HRI. An effective Gabor filter was utilized alongside a set of morphological and convolutional filters to lessen the noise and the light dependence of the image acquired by the robot [14]. Wenbai et al. realized human's gesture recognition and imitation by humanoid robot NAO, Based on Kinect platform [15]. Mao et al. proposed a real-time emotion recognition approach based on both 2D and 3D facial expression features captured by Kinect sensors [16]. Chun Fui Liew et al. evaluated five most commonly used feature spaces with seven classification methods to identify the most effective features for facial expression recognition [17].

In this paper, we try to address the problem of facial expression recognition and real-time imitation by a social humanoid using artificial neural networks. The rest of the paper is organized as follows: Section II outlines characteristics description of the social humanoid, Section III describes the procedure of key point's extraction, Neural Networks development and social robot facial gesture imitation, and section IV presents the experimental resultsand the last section provides conclusions.

978-1-5090-3222-8/16/$31.00 ©2016 IEEE

Fig. 1. R50 (Alice) robot.

II. R50 (ALICE) ROBOT

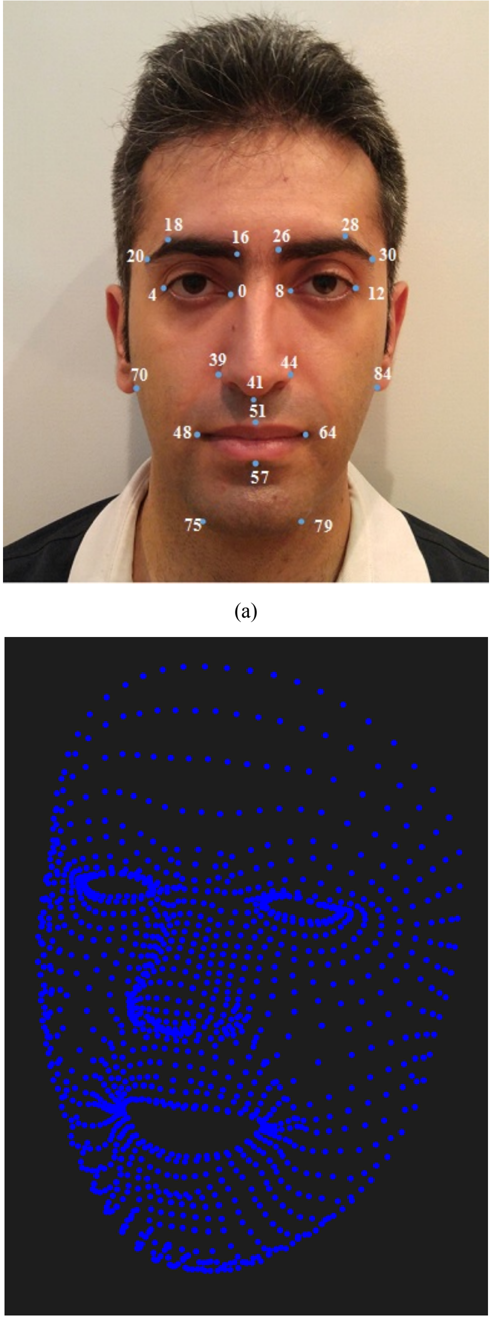
For the experimental study, the social humanoid called Alice is used. The robot Alice is a humanoid robot with a full range of facial muscles allowing a broad display of expressions including: smile, frown, blink, angry, surprised, happy and sad. The robot Alice is made by Robokind Co. with total 32 degrees of freedom (DOF). It has 11 *DOFs* in the head, where 8 *DOFs* are for facial expression (Brows, Eyelids, Eyes pitch, Left eye yaw, Right eye yaw, Left smile, Right smile and Jaw) and 3 remaining DOFs are in her neck (Neck yaw, Neck roll and Neck pitch). The robot Alice is shown in the figure 1.

III. ROBOT FACIAL EXPRESSION IMITATION USING KINECT

A. *Extracting Key Points and Angles*

In this approach, the sensing device is the Kinect, the Kinect sensor is a line of motion sensing input devices that produced by Microsoft Company. Using Microsoft Software Development Kit (SDK) 2D face detection, Kinect can detect approximately 100 keypoints in the face. In addition, Kinect can detect more than 1300 facial keypoints by using SDK 3D HD-face detection code (Fig.2 (b)). The 22 facial keypoints from 2D SDK face detection code, which are useable in this study are shown in Figure.2(a). These keypoints indexes are: 0, 4, 8, 12, 16, 18, 20, 26, 28, 30, 39, 41، 42, 44, 48, 51, 54, 57, 70, 75, 79 and 84. By running 3D HD-face code together with Kinect, we are able to extract 17 animation units, which are generated by SDK based on the distances between the facial keypoints. We manipulate the code to extract these units and four quaternion angles from Kinect, the 17 facial units are labeled as: Jaw Open, Lip Pucker, Jaw Slide, Lip Stretcher, Right Lip Stretcher Left, Lip Corner Puller Left, Lip Corner Puller Right, Lip Corner Depressor Left, Lip Corner Depressor Right, Left Cheek Puff, Right Cheek Puff, Left Eye Closed, Right Eye Closed, Right Eyebrow Lower, Left Eyebrow Lower, Lower Lip Depressor Left, Lower Lip Depressor Right. We use these 17 data as an array for next processes such as an input array for the classification algorithm of our approach.

Kinect extracts quaternion angles to avoid gimbal lock. To actuate head and neck of the robot to imitate the movement of the head of the human we need to obtain Euler Angles. Though, having Euler angles, Quaternion Euler transfers are used; and Roll, Pitch and Yaw for the head of the humanoid are extracted.



(a)

(b)

Fig. 2. (a) 2D detected facial keypoints, (b) 3D detected facial keypoints.

In this case, we have two situations; the first one is the exact imitation that we don't have any changes in the extracted Euler angles. But in the other one, which is named as the mirror imitation, we need to multiple -1 to the extracted Roll and Yaw angles. The first situation is used in indirect interaction with patient or a kid; but the second one is used in face to face interaction with patient or user.

B. *Facial Expression recognition*

In this section the human facial expressions are recognized and classified in 6 categories, including: happy, sad, angry, surprised, disgusted and neutral. The method used to recognize facial expressions is Artificial Neural Networks (ANN). We utilize ANN to classify the output data array from Kinect and to create training dataset for the Neural Networks.

1) Creating training dataset: Kinect output array are captured and collected for each expression. The training dataset include 10 persons, 6 males and 4 females (Fig. 3). All persons' Kinect data output was captured at least for 3 seconds with 30 frame per second rate for each facial expression with different head orientation. Though, each person has a dataset that each frame of data is an array with 17 members.

2) Finding optimum network for real-time recognition: In this stage, we are going to find optimom Multi-Layer Perceptron (MLP) neural networks to decrease recognition time for real-time reaction of the robot. In the beginning, for each expression, the network has 17 inputs (Kinect output) and 6 outputs. we created 3 separated fully-connected MLP networks with 7 layers that trained by 3 different methods such as: Gradient descent back propagation (GD), Gradient descent with momentum and adaptive learning rate back propagation (GDMA) and Levenberg-Marquardt back propagation (LM). Then, we change the numbers of neurons in each layer or the arrangement of the network in heuristic way. we decrease network size layer by layer and train the new networks to reach the one that has the minimum layers with desirable accuracy in expressions recognition.

a) Training Networks: To train these three networks (GD, GDMA and LM) the training parameters are set as table I. the GD method learning procedure is slower and more time consuming than two other methods. We set minimum performance gradient (MPG) for the network with GD training method equal to 10^{-6} and it takes 10^{7} epochs to reach this gradient. Training in LM and GDMA methods takes less than 500000 epochs to reach MPG.

b) Testing Networks: In all methods 15% of the dataset is used for testing networks randomly. Minimum error in classification of the test data belongs to GD methods but LM is more accurate in output values so that the output layer values are closer to 0 or 1.

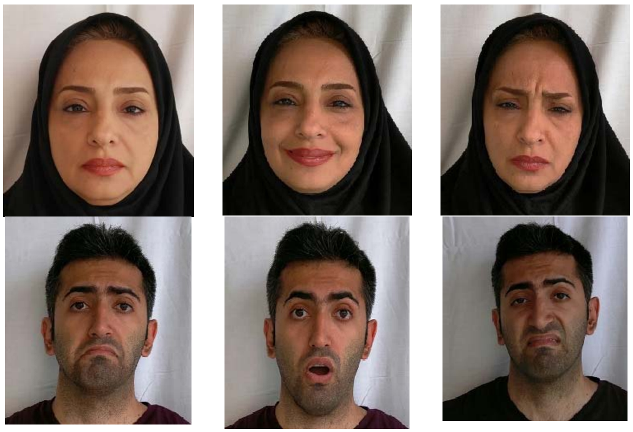


Fig 3. The training dataset

c) Eliminating layers: The next step is declining the hidden layers of the networks and go back to step (a) and afterward test the new networks to examine the accuracy.

After examining more than fifty networks we find the network with minimum hidden layers that is more accurate than the others. This neural network has three layers and number of neurons in each layer is six for input layer, ten in second hidden layer and six in output layer; and the training method is LM (see Fig. 4). The neural network facial expression classification results are shown in Table II.

TABLE I. TRAINING PARAMETERS

Neural Network parameters	***GD***	***GDMA***	***LM***
Learning rate	0.03	0.03	-
Ratio to increase learning rate	-	1.01	-
Ratio to decrease learning rate	-	0.8	-
Maximum performance increase	-	1.01	-
Momentum constant	-	0.95	-
Minimum performance gradient	10^{-9}	10^{-9}	10^{-6}
Initial μ	-	-	0.001
μ decrease factor	-	-	0.2
μ increase factor	-	-	7
Maximum μ	-	-	10^{10}

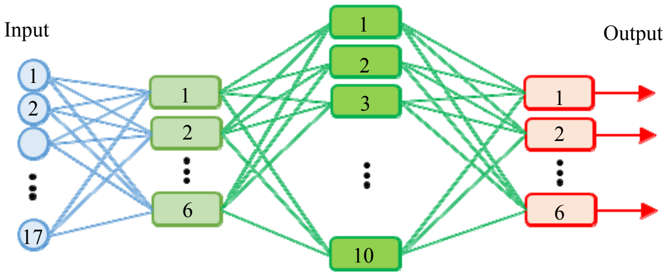


Fig.4. Final Neural network

TABLE II. CLASSIFICATION RESULT

	Happy	Sad	Angry	Surprised	Disgust	Neutral
Happy	98	0.7	0.1	0.1	0.1	1
Sad	1.5	87	1.8	0.7	3.5	5.5
Angry	0.9	1.7	95.3	0.1	0.5	1.5
Surprised	0.1	0.2	0.7	94.8	4	0.2
Disgust	0.2	2.3	0.7	4.8	91	0.5
Neutral	3.2	5.1	1.1	0.1	1.5	89

Besides this 3-layer network we find a 2-layer network with lower accuracy that is trained with GD method. This network has ten neurons in first layer and six neurons in output layer. The network is noticeable because regardless of the accuracy; heuristically, we recognized this network can detect some combined expression such as a gesture with angry eyebrows and happy lip and mouth. The classification output of the 2-layer network for an example of this expression is happy = 0.97, angry = 0.94 and the other outputs are close to zero.

3) Applying mathematical equivalent of neural network to compiler: In this stage we turn the offline recognition to real-time recognition. Due to this procedure, the acquired neurons weights and biases in training section should be applied to the compiler. The weights and biases are implement in the neurons transfer function layer to layer as follows;

$$y_i = F\left(w_{ij} \times x_j + b_i\right) \tag{4}$$

where *y, w, x, b, i, j,F* are neuron output,neuron weight value,neron input,neuron bias value, number of neuron, number of neurnon input and neuron transfer function Respectively. In this study all neurons have tan-sigmoid transfer function as:

$$tansig\ (n)= 2/(1+exp(-2n))\ \text{-}1 \tag{5}$$

Where *n* is an input of transfer function.

C. Social Humanoid Imitation

The final step is making connection between Kinect sensor and the robot Alice to complete the imitation process. As we mentioned, Alice social humanoid works with Java NetBeans and Kinect recognition processes takes place in C#. Though, for connecting compilers and sending data a socket code is created between these compilers. We have generated six face gestures for the robot which are equivalent to the human facial expressions. As a result, the expressions detected by the neural network and the humanoid imitate the proper gestures. Furthermore, the Euler angles that we calculated previously have been send with the socket; so robot can imitate the head movement of the user.

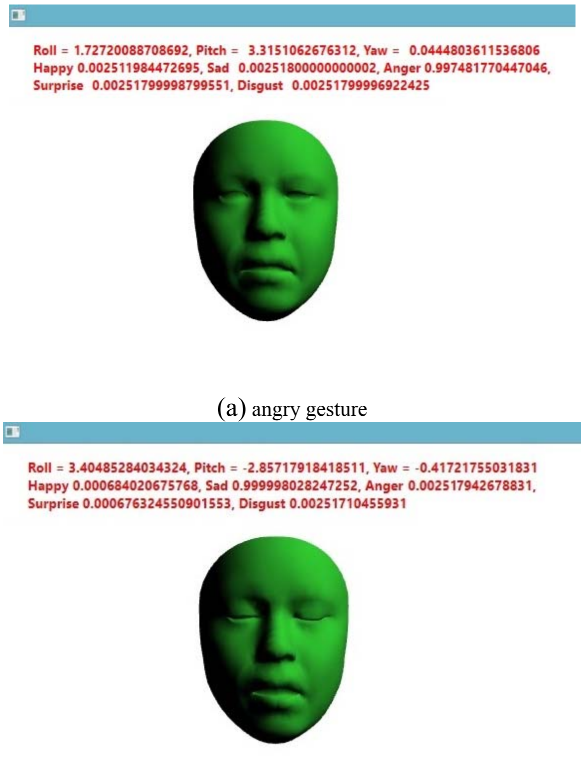
IV. Experimental Results

To evaluate the real-time facial expression and head orientation recognition, the method is simulated and then it implemented on the social humanoid Alice. The neural network that was applied in C# is run with the Kinect sensor to demonstrate the accuracy of the method. Figure.5 exhibit the snapshots of two gestures which are recognized by the ANN, the angry gesture (Fig.5 (a)) and the sad gesture (Fig.5(b)), and the outputs values for each detected gestures are reported in table III. As it is exposed in snapshots of the simulation (Fig.6.) orientation of users head is shown with Euler angles in first line (angles are in degree) and classification results of facial expressions are shown in other lines as a degree of membership of detected gesture in each basic facial gestures classes. Outputs for the images of the snapshot are shown in table III. As it is shown in Fig.5 and table III for three different orientations of users' head, (a) frontal face, (b) looking to the right and (c) looking to the left, facial expression recognition output values indicate the recognition of happy facial gesture in user for all orientations. Extracted Euler angles exposed in Fig.6. (a), (b) and (c) show the correctness of our method.

TABLE III. Outputs value in snapshot

Outputs	Output Value				
	Fig.5 (a)	Fig.5 (b)	Fig.6 (a)	Fig.6 (b)	Fig.6 (c)
Roll (degree)	1.727	3.40	-10.59	-2.73	0.35
Pitch (degree)	3.31	-2.85	-.3.91	8.32	-7.03
Yaw (degree)	0.04	-0.41	-3.81	42.65	-45.4
Happy	0.0025	0.0006	0.9973	0.9954	0.9955
Sad	0.0025	0.9999	0.0025	0.0022	0.0022
Angry	0.997	0.0025	0.0025	0.0024	0.0024
Surprised	0.0025	0.0006	0.0025	0.0025	0.0025
Disgust	0.0025	0.0025	0.0025	0.0025	0.0025

At last the Java Netbeans codes and C# codes are run together with Kinect to move the robot. The robot imitation indicated the appropriate performance of Neural Network. As it is shown in snapshots of experimental results, the robot imitates the users' facial gesture from neutral to surprised (Fig.7.). Figure.7 (a) shows the neutral facial expression imitation, (b) shows surprised facial expression imitation and (c) shows the frontal view of Alice in middle of surprised facial expression imitation. The Alice response time was desirable and it was less than 0.5 second, which is satisfying our goal as a real-time reaction.



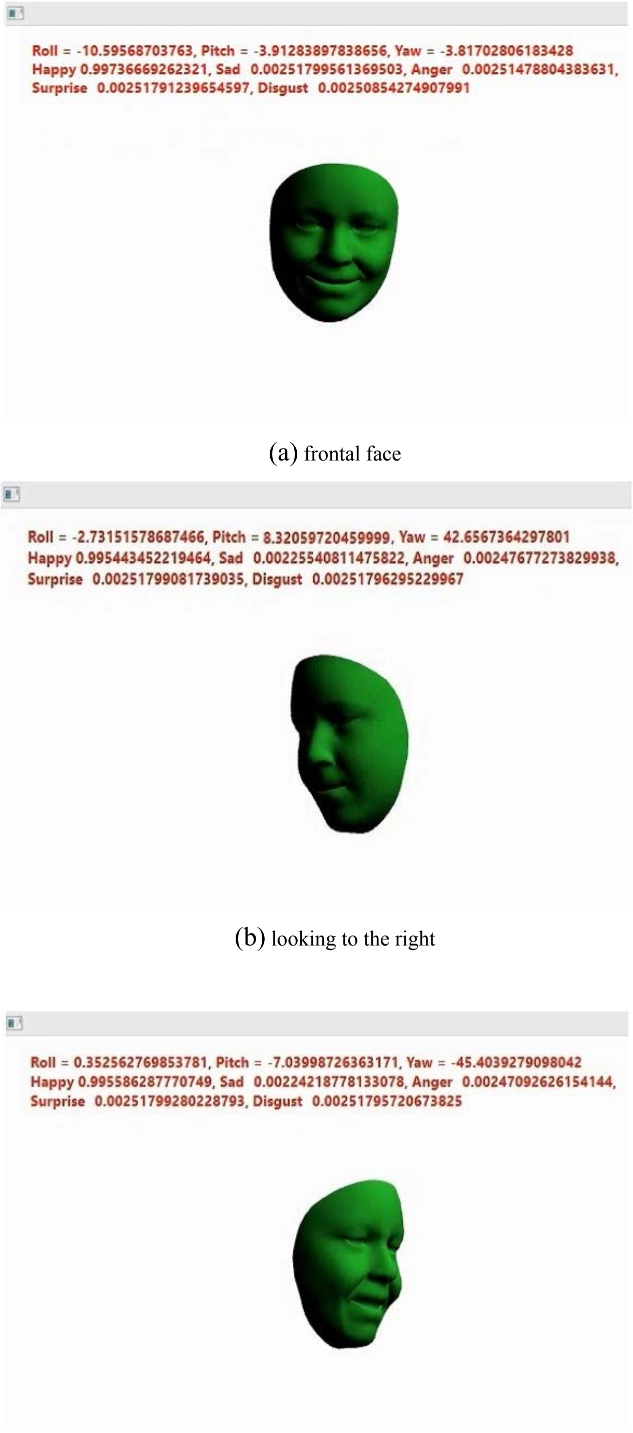
```
Roll = 1.72720088708692, Pitch =  3.3151062676312, Yaw =  0.0444803611536806
Happy 0.002511984472695, Sad  0.00251800000000002, Anger 0.997481770447046,
Surprise  0.00251799998799551, Disgust  0.00251799996922425
```

(a) angry gesture

```
Roll = 3.40485284034324, Pitch = -2.85717918418511, Yaw = -0.41721755031831
Happy 0.000684020675768, Sad 0.999998028247252, Anger 0.002517942678831,
Surprise 0.000676324550901553, Disgust 0.00251710455931
```

(b) sad gesture

Fig.5. Snapshots of simulation results for two detected facial gestures.



```
Roll = -10.59568703763, Pitch = -3.91283897838656, Yaw = -3.81702806183428
Happy 0.99736669262321, Sad  0.00251799561369503, Anger  0.00251478804383631,
Surprise  0.00251791239654597, Disgust  0.00250854274907991
```

(a) frontal face

```
Roll = -2.73151578687466, Pitch = 8.32059720459999, Yaw = 42.6567364297801
Happy 0.995443452219464, Sad  0.00225540811475822, Anger  0.00247677273829938,
Surprise  0.00251799081739035, Disgust  0.00251796295229967
```

(b) looking to the right

```
Roll = 0.352562769853781, Pitch = -7.03998726363171, Yaw = -45.4039279098042
Happy 0.995586287770749, Sad  0.00224218778133078, Anger  0.00247092626154144,
Surprise  0.00251799280228793, Disgust  0.00251795720673825
```

(c) looking to the left

Fig.6. Snapshots of simulation result and different orientations of happy facial expressions.

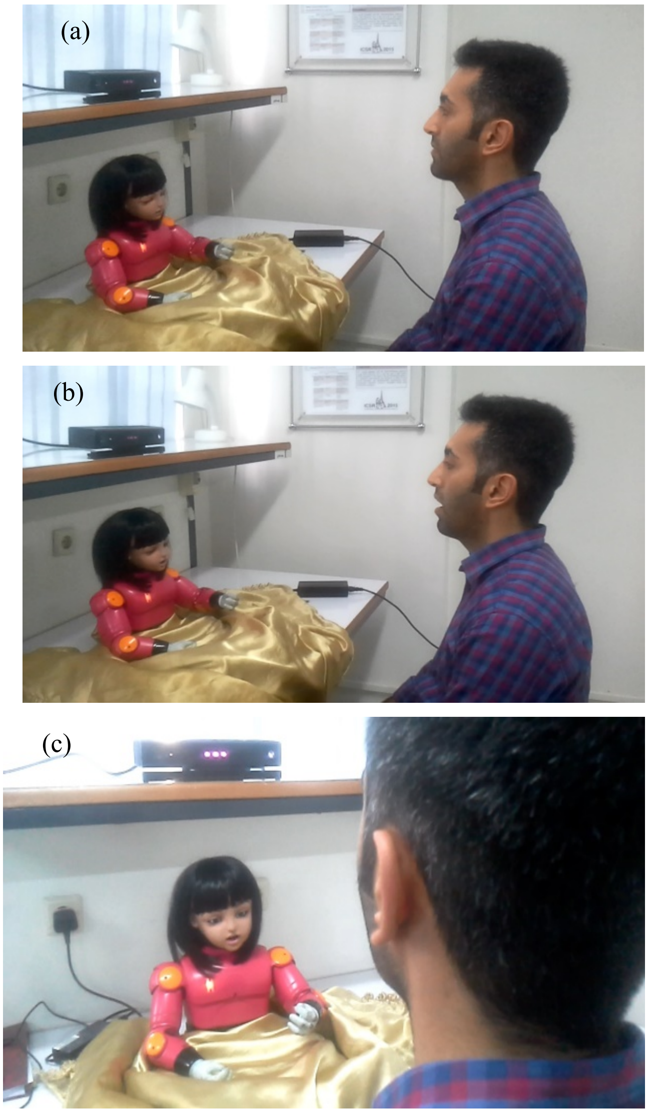


Fig.7. Snapshots of the robot imitation, (a) neutral facial expression imitation, (b) surprised facial expression imitation and (c) frontal view of Alice in middle of surprised facial expression imitation

V. Conclusion

This paper contributes to the real-time facial expression imitation of a human, by the humanoid robot, Alice. The key points of the user are extracted with the Kinect sensor and SDK. The human facial expressions are recognized and classified using Artificial Neural Networks. The neural network has been developed with the minimum layers. Kinect output array are captured and collected for each expression and training dataset is created. Then, optimum network for real-time recognition is found and trained. Then, two compilers are connected to complete the imitation process. Simulations and experimental results showed a satisfying reaction time and appropriate accuracy in gesture recognition and imitation by the social humanoid Alice.

REFERENCES

[1] M. Alemi, A. Meghdari, and M. Ghazisaedy, "The impact of social robotics on L2 learners' anxiety and attitude in english vocabulary acquisition," International Journal of Social Robotics, vol. 7, pp. 523-535, 2015.

[2] A. Taheri, M. Alemi, A. Meghdari, H. Pouretemad, N. M. Basiri, and P. Poorgoldooz, "Impact of Humanoid Social Robots on Treatment of a Pair of Iranian Autistic Twins," in *Social Robotics*, ed: Springer, 2015, pp. 623-632.

[3] A. Meghdari, M. Alemi, M .Ghazisaedy, A. Taheri, A. Karimian, and M. Zandvakili, "Applying robots as teaching assistant in EFL classes at Iranian middle-schools," in *CD Proc. of the Int. Conf. on Education and Modern Educational Technologies (EMET-2013)*, 2013.

[4] E. P. Ijjina and C. K. Mohan, "Facial Expression Recognition Using Kinect Depth Sensor and Convolutional Neural Networks," in *Machine Learning and Applications (ICMLA), 2014 13th International Conference on*, 2014, pp. 392-396

[5] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry,* vol. 32, pp. 88-106, 1969.

[6] A. Taheri, M. Alemi, A. Meghdari, H. PourEtemad, and N. M. Basiri, "Social robots as assistants for autism therapy in Iran: Research in progress," in *Robotics and Mechatronics (ICRoM), 2014 Second RSI/ISM International Conference on*, 2014, pp. 760-766.

[7] A. Paiva, J. Dias, D. Sobral, R. Aylett, P. Sobreperez, S. Woods*, et al.*, "Caring for agents and agents that care: Building empathic relations with synthetic agents," in *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*, 2004, pp. 194-201.

[8] M. Siegel, C. Breazeal, and M. I. Norton, "Persuasive robotics: The influence of robot gender on human behavior," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 2563-2568.

[9] S. S. Ge, C. Wang, and C. C. Hang, "Facial expression imitation in human robot interaction," in *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 2008, pp. 213-218.

[10] S. DiPaola, J. Chan, and A. Arya, "Simulating face to face collaboration for interactive learning systems," 2005.

[11] I. Song, H.-J. Kim, and P. B. Jeon, "Deep learning for real-time robust facial expression recognition on a smartphone," in *2014 IEEE International Conference on Consumer Electronics (ICCE)*, 2014, pp. 564-567.

[12] S. E. Kahou, C. Pal, X. Bouthillier, P. Froumenty, Ç. Gülçehre, R. Memisevic*, et al.*, "Combining modality specific deep neural networks for emotion recognition in video," in *Proceedings of the 15th ACM on International conference on multimodal interaction*, 2013, pp. 543-550.

[13] P. Barros, C. Weber, and S. Wermter, "Emotional expression recognition with a cross-channel convolutional neural network for human-robot interaction," in *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*, 2015, pp. 582-587.

[14] F. Cid, J. A. Prado, P. Bustos, and P. Núnez, "A real time and robust facial expression recognition and imitation approach for affective human-robot interaction using gabor filtering," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 2188-2193

[15] C. Wenbai, W. Xibao, W. Sai, and G. Hui, "Human's Gesture Recognition and Imitation Based on Robot NAO," *International Journal of Signal Processing, Image Processing and Pattern Recognition,* vol. 8, pp. 259-270, 2015.

[16] Q.-r. Mao, X.-y. Pan, Y.-z. Zhan, and X.-j. Shen, "Using Kinect for real-time emotion recognition via facial expressions," *Frontiers of Information Technology & Electronic Engineering,* vol. 16, pp. 272-282, 2015

[17] C. F. Liew and T. Yairi, "A comparison study of feature spaces and classification methods for facial expression recognition," in *Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on*, 2013, pp. 1294-129.