Roboticon: A realistic platform to imitate facial expressions

Conference Paper · December 2016

DOI: 10.1109/ICECE.2016.7853932

Amit Kumar Kundu
Bangladesh University of Engineering and Technology

Md Didarul Islam
University of Nebraska–Lincoln

Roboticon: A Realistic Platform to Imitate Facial Expressions

Amit Kumar Kundu*, Md. Didarul Islam* and S. M. Mahbubur Rahman †
*Department of Electrical and Electronic Engineering, Uttara University, Bangladesh
†Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Bangladesh
Email: amit31416@gmail.com, s102bueeet@yahoo.com, mahbubur@eee.buet.ac.bd

***Abstract*—This paper presents a concept that quickly detects the facial expressions of humans with a sufficient accuracy and imitates the expressions remotely in terms of robot-based mimics. In traditional online communication of expressions, it is not certain that the sender remains in the same state of expression, when the state is actually received in the remote side. In addition, the expression received in the remote side is mimicked by special visual characters commonly known as emoticons. The approach presented in this paper addresses the issues of real-time detection, communication, and imitation of expressions in order to solve the existing shortcomings. In particular, a computationally efficient intensity threshold-based algorithm is adopted to detect facial expressions from video signals. Because of such a simplicity, the proposed system ensures fast detection of an expression, which is quite convenient for real-time communication. At the same time, the system copies the expression on the face of a humanoid robot at the receiving end. It is expected that this novel approach, which is referred to as Roboticon will play a key role to imitate facial expressions in real-time.**

Keywords*—Action units, expression detection, expression imitation, interfacing, Viola-Jones algorithm.

I. Introduction

Facial expression is considered to be the most effective nonverbal way for human beings to express their emotions and interact with others. Emoticons and video calls are already being used for expressive communications. But a more effective representation would be a holographic or robotic face that can actually mimic someone's expressions. In this research work, focus has been given on copying sender's expression on a robotic face instead of emulating expression on a holographic one. Therefore, it is necessary for an expression copying robot to recognize and imitate human facial expressions as close as possible [1].

Many researchers have attempted to automatically recognize facial expressions of human beings. But understanding and imitating the human emotions appropriately have been remained a challenge in today's robotics while researchers are trying to make the robots with high level of sophistications [2]. For example, in [3], facial expression is detected utilizing twenty two facial muscle tensions. The 2D Gaussian-Hermite moments of facial distortions from the neutral images are used for recognizing human expressions in [4]. In order to express various emotions on a robot head; arm type antennae, emoticon eyes and exaggerating jaws and checks are used in [5].

In this paper, a simple, fast but efficient algorithm is proposed for automatic expression recognition which can be easily implemented in real time communication. The research scope of the paper consists of three expressions: happiness, sadness and surprise. Simple algorithm is used here for fast and real time implementation. At the receiving end a realistic approach of human expression imitation is attempted.

The organization of the rest of the paper can be summarized as follows. In Section II, methodology of the system is described. Section III presents the experimentation part where the whole idea is implemented practically. Finally, in Section III, conclusion is drawn and some future directions of works are proposed.

II. Methodology

The proposed facial emotion detection method comprises of three major parts which are expression detection, expression transmission and expression imitation as depicted in Fig. 1. In the first segment, images are acquired continuously. Then, a simple threshold based scheme is proposed to detect facial emotions such as: happiness, sadness or surprise. These expressions are detected based on RGB values in the informative portion of test images and after that by comparing them with their corresponding RGB values of neutral images of the same test subject. Later, depending on the expression of consecutive frames a decision making algorithm is proposed to confirm the facial expression in case of a falsely detected expression. The second step involves the transmission part from one computer to another computer. The detected expression is sent to another computer for the purpose of imitating that expression at the receiving end. The last part is the expression imitation process in which corresponding expression is imitated in a robot face with the help of LEDs and Arduino interfacing.

A. Image Acquisition

In order to detect facial expression, real time image of a human face is needed to be acquired. To serve that purpose, CCD camera is used to acquire continuous images. Fig. 2 shows an example of continuous image acquisition of a test subject. Then, individual frames are extracted and fed to the next step to recognize expression for each one separately.

B. Expression Detection for Individual Frame

To avoid false expression detection, it is required to detect expression of individual frames and ensure that this is the correct expression. At first, Viola-Jones algorithm is implemented to detect face parts such as: mouth, eye-pair and nose of a subject. This algorithm is widely used for many cascade

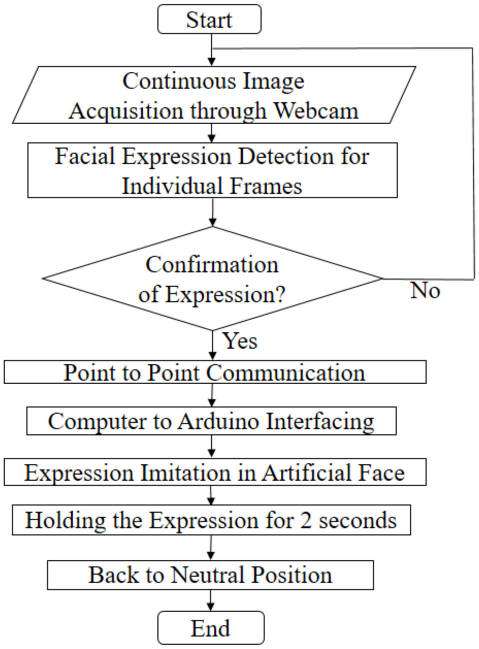

Fig. 1: Overview of system architecture.

object detection applications [6]. It is seen that for some subjects, multiple mouths or eye-pairs are detected considering environmental conditions (e.g., lighting, background, photo quality). Even though the number of multiple detection is too few, this multiple detected mouths or eye-pairs are on the opposite sides (up and down) of the nose. Hence, to remove this redundancy, coordinates of detected mouths or eye-pairs are compared with respect to coordinate of nose and falsely detected mouth or eye-pair is removed.

Next, from separate face parts, expressions are detected based on simple RGB values. For instance, smiling varies from person to person. But for simplicity, it is assumed that a person will reveal his teeth at the time of smiling. So, based on the RGB values of teeth, the areas of teeth are calculated and compared with his neutral image. If the teeth area of an image exceeds a certain threshold value (teeth area of neutral face), then the expression can be considered as happy. Similarly, it is expected that for a surprised face, the eye-ball area will increase with respect to neutral eye. Utilizing this technique, a surprised expression is detected. Last of all, for sad expression, it is assumed that the corner of the lips will get lower than the neutral lips corner. Therefore, distance of lips corner from nose is measured and compared with respect to neutral expression. If this distance gets increased by a certain amount then the image is considered as sad expression.

C. Decision Making

After detecting expression for an individual frame, decision of consecutive frames are compared to confirm which expression is shown by the test subject. If the detected expression is neutral, the system will continue to check the next frame. But in case of any other expression (except neutral), if that expression is detected for five or more consecutive frames, that expression is saved temporarily. Otherwise (non-neutral expression's number of frame is less than five), the system does not save any expression temporarily rather searches for next expression. This step helps to remove falsely detected expressions as any expression cannot stay in progress only for one or two image frames. After that, when expression is over, it is expected that, neutral expression will be back in the next frames. Therefore, this feature is incorporated in the algorithm that if three consecutive frames are detected as neutral (after a temporarily saved non-neutral expression), the temporarily saved expression will be confirmed and will indicate the end of expression. Later, this expression is sent to remote end for imitation. Once the whole process is completed the detection algorithm will search for new expressions.

TABLE I: ACTION UNIT NUMBER OF DIFFERENT EXPRESSIONS

Emotion	Action Units	LED Color
Happiness	6+12	Green
Sadness	1+4+15	Red
Surprise	1+2+5+26	Yellow

TABLE II: NAME OF EACH ACTION UNIT

AU Number	FACS Name
1	Inner Brow Raiser
2	Outer Brow Raiser
4	Brow Lowerer
5	Upper Lid Raiser
6	Cheek Raiser
12	Lip Corner Puller
15	Lip Corner Depressor
26	Jaw Drop

D. Point to Point Communication

In this research work, sending and receiving points (computers in our case) are connected via physical links and communicate through a suitable program (MATLAB is used for implementation). The detected expressions from sending end computer is transmitted to the receiving end computer online.

E. Expression Imitation

Once an expression is received by the receiving end computer, that expression is required to be emulated on a humanoid robotic face. Suitably placed bidirectional motors (in the robotic face) will rotate in one direction to activate certain action units and after a specific time, will rotate in opposite direction to show the neutral expression again. Facial action unit system (FACS) is a system to express the facial muscle movement during any expression given by a person [7]. Table I shows the action unit numbers that get activated during any expression and Table II shows the name of action units and their corresponding numbers.

III. EXPERIMENTATION

For experimentation purpose, 3 test subjects (of 3 different ages) are chosen for expression detection. Although there are six basic classes of expressions, this paper deals with these three expressions: happiness, sadness and surprise [8]. These particular three emotions were chosen as their clusters are most distant with respect to the neutral expression (see for example in [4]). The CCD camera connected with the sender's

Fig. 2: Continuous image acquisition (oldest to newest frame from left to right).

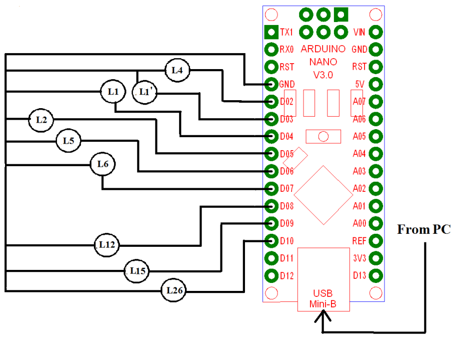

Fig. 3: Arduino connection diagram.

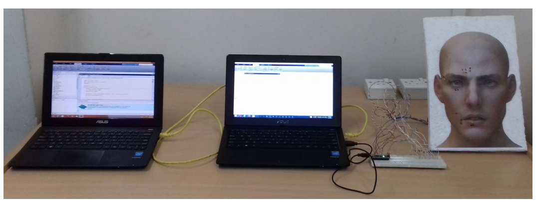

Fig. 4: Overall setup (from left to right: sender PC, physical link, receiver PC, Arduino, Robotic face).

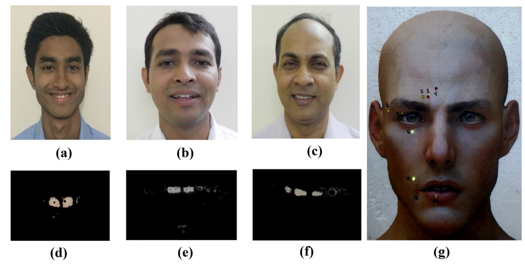

Fig. 5: Happy expression detection; (a), (b), (c) Happy expressions of the subjects; (d), (e), (f) Teeth areas of happy expressions; (g) Happy expression imitation on robotic face.

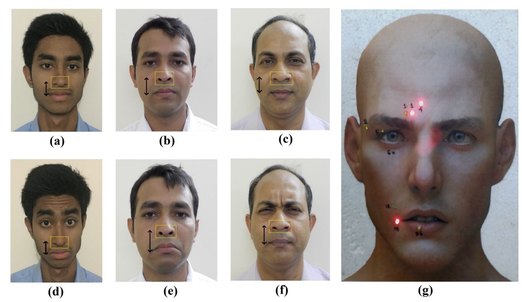

Fig. 6: Sad expression detection; (a), (b), (c) Sad expressions of the subjects (Nose to lip corner distance is shown); (d), (e), (f) Lip corner depression in sad expressions (Nose to lip corner distance increased); (g) Sad expression imitation on robotic face.

computer is used for continuous image acquisition at a rate of 5fps. After that, test subject's expressions are detected from continuous images, transmitted to receiving end computer and implemented there on a robotic face with the help of Arduino interfacing. Fig. 3 and Fig. 4, respectively, show the Arduino LED connection diagram and the setup of overall system. To detect expression from individual images, at first, Viola-Jones algorithm is used to detect face parts such as: mouth, eye-pair and nose. Then, individual expressions from images are detected using simple threshold based algorithm utilizing the RGB values of different face parts.

After detecting mouth of a subject, teeth is distinguished from its RGB values. Measurement shows that RGB values lie between 80 to 160 for teeth pixels for the considered subjects. Fig. 5((a)-(f)) shows the happy faces of the subjects and their corresponding detected teeth areas. These teeth areas are compared with the teeth areas of neutral faces. It is observed that in happy expression the teeth area is around 7 times greater. So, if the teeth area in an image is found 7 times or more than the teeth area in the neutral expression, then the expression is considered as happy. For a happy expression, action units 6 and 12 get activated which is shown with the help of green LEDs in Fig. 5(g). However, it is to be noted that teeth area does not increase that much for other expressions.

It is obvious that, for sad expression, the lips corner depresses. Experimentation shows that RGB values of lip corner lies between 25 to 35. Distance between nose level to the lip corner is measured. If the distance increases compared to their neutral value then, the subject is considered to be a sad. Fig. 6((a)-(f)) shows the neutral faces and respective changes in sad faces of all subjects. For a sad expression, action units 1, 4 and 15 get activated which is shown with the help of red LEDs in Fig. 6(g).

Similarly to detect surprised expressions, it is assumed that a persons eyeball will get bigger when he/she is surprised. Therefore, after properly detecting the eye-pair, the area of eyeball is detected using RGB values. Experimentation shows that RGB values lie between 90 to 150 for eyeball pixels. Fig 7((a)-(f)) shows the surprised faces of the subjects and their corresponding detected eyeball areas. After detecting the surprised eyeball areas, they are compared with the neutral eyeball areas and it is seen that in surprised images the eyeball areas is double or more than the eyeball area in the neutral image. This technique is used to detect surprised expressions. Action units 1, 2, 5 and 26 get activated for surprised expression, which is imitated in robotic face by

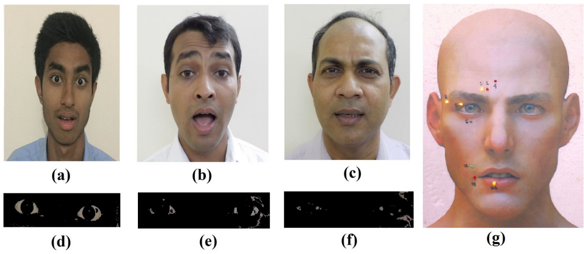

Fig. 7: Surprised expression detection; (a), (b), (c) Surprised expressions of the subjects; (d), (e), (f) Eyeball areas of surprised expressions; (g) Surprised expression imitation on robotic face.

yellow LEDs in Fig. 7(g). However, it is to be noted that for other considered expressions eyeball area remains same as neutral expression.

Once an expression is confirmed by the decision making algorithm, it is sent to the receiving end computer for imitation purpose. This transmission of expression is done via TCP/IP based socket communication with MATLAB [9]. The IP address of both host and server computer is required for this purpose.

After receiving the expression from the sending computer, the expression is intended to perform by bidirectional motors. But for better visual representation, LEDs are used instead of bidirectional motors. Three different color LEDs are lit for 2 seconds to show the activation of certain action units and thus imitating different expressions as explained in Table I. The LEDs are operated using arduino interfacing [10].

IV. Conclusion

In this paper, a simple but effective expression detection scheme in RGB color space has been proposed. The RGB values of lips, teeth, and eyeball have been found to be very effective for differentiating among different expressions. The simple threshold based scheme makes the process of expression detection and communication faster and involves less computation. In the experimentations on test subjects of different ages it is observed that the proposed method is capable of detecting almost all the expressions correctly with a rare exception. Therefore, it can be concluded that the proposed system can serve as an effective way to transmit real-time expressions on-line. The Arduino-based LED outputs on the FACS open the door of robotic imitation of expressions in the receiving end. In future, the system can be extended to communicate wirelessly using Bluetooth module (e.g., BLE v4.0). Moreover, other emotional expressions such as anger and disgust can be incorporated to the system along with happiness, sadness, and surprise. At the same time, the imitation of expressions can be confirmed as 'seen', when the receiving end uses motorized units on facial parts of robots.

References

[1] Y. Yang, S. S. Ge, T. H. Lee, and C. Wang, "Facial expression recognition and tracking for intelligent human-robot interaction," *Intelligent Service Robotics*, vol. 1, no. 2, pp. 143–157, 2008.

[2] S. S. Ge, "Social robotics: Integrating advances in engineering and computer science," in *Proceedings of Electrical Engineering/ Electronics, Computer, Telecommunications and Information Technology International Conference*, 2007, pp. xvii–xxvi.

[3] S. S. Ge, C. Wang, and C. C. Hang, "Facial expression imitation in human robot interaction," in *Proceedings of 17th IEEE International Symposium on Robot and Human Interactive Communication*, Munich, Germany, 2008, pp. 213–218.

[4] S. M. Imran, S. M. Mahbubur Rahman, and D. Hatzinakos, "Differential components of discriminative 2D Gaussian-Hermite moments for recognition of facial expressions," *Pattern Recognition*, vol. 56, pp. 100–115, 2016.

[5] H. Song, Y. M. Kim, J. C. Park, C. H. Kim, and D. S. Kwon, "Design of a robot head for emotional expression: Eeex," in *Proceedings of 17th IEEE International Symposium on Robot and Human Interactive Communication*, Munich, Germany, 2008, pp. 207–212.

[6] Y. Muttu and H. G. Virani, "Effective face detection, feature extraction neural network based approaches for facial expression recognition," in *Proceedings of International Conference on Information Processing (ICIP)*, Pune, India, 2015, pp. 102–107.

[7] FACS- Facial action coding system. [Online]. Available: https://www.cs.cmu.edu/~face/facs.htm [Accessed on August 30, 2016].

[8] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[9] Robotics, teaching and learning. [Online]. Available: http://www.lucidarme.me/?p=4087 [Accessed on August 30, 2016].

[10] Arduino. [Online]. Available: https://www.arduino.cc [Accessed on August 30, 2016].