Real-Time Emotional Expression Generation by Humanoid Robot

Master Prince
Assistant Professor, Department of Computer Science
College of Computer, Qassim University, Buraidah, 51452, Al Qasim, Saudi Arabia

***Abstract*—Emotion integrates different aspects of a person, including mood (current emotional state), personality, voice or speech, color around the eyes, and facial organs' movement. We are considering the mood because a person's current emotional state must always affect upcoming emotions. So behind an emotion, all these parameters are involved, and a human being can easily recognize it by seeing that face even if more than one person is there, so for the robot to make human-like emotion, all these parameters have to be considered to imitate artificial facial expression against that emotion. Most researchers working in this area still find difficulties in determining exact emotion by the robot because facial information is not always available, especially when interacting with a group of people and mimicking exact emotion that the user can effectively recognise. In our study, the loud most speeches among the people sensed by the robot and color around eyes are considered to cope with these issues. Another issue is the rise time and fall time of emotional intensity. In other words, how long should the robot keep an emotion here? An experimental approach is applied to get these values. The proposed method used an emotional speech database to recognize the human emotion using convunational neural network (CNN) and RGB patterns to mimic the emotion, which simulates an improved humanoid robot that can express emotion like human beings and give real-time responses to the user or group of users that can make more effective Human-Robot Interaction (HRI).**

Keywords—Artificial facial expression; emotional speech database; convunational neural network; RGB pattern; humanoid robot; human-robot interaction

I. INTRODUCTION

The human face is very special in different aspects; one of those aspects is expressing emotion. By expressing emotion, human beings express their feelings, and others can easily understand the feeling and respond as well [Y. Yang et al., (2007)]. However, when we talk about HRI, it becomes challenging for the humanoid robot to determine the exact emotion expressed by the person (human) who is interacting at a particular moment, especially when interacting with a group of people and when emotion is not evident with the face because according to psychologist numerous types of expression can be produced by a human. Moreover, we have only seven recognized expressions: Natural, Happy, Sad, Anger, Surprise, Fear and Disgust.

Robotics has become a very emerging area in today's world; it plays a significant role in various fields like medical science, military applications, home appliances, education, and many more. In recent years, a popular research area in robotics has been developing intelligent robots that can interact with people as companions rather than machines. To interact with a humanoid robot, HRI is very important; studies of human-robot interaction will be improved by automated emotion interpretation. A humanoid robot must be able to understand the person's actual emotional state at a particular instance.

Here the proposed method used our previous emotion recognition method that represents the intensities of the emotions instead of emotion. Once the intensity of the emotion was known, the main goal was to determine fusion weight for each primary emotion based on all those parameters, which include mood, personality, and intensities of the recognized emotional states employing fuzzy Kohonen clustering network (FKCN), which give us a smooth variation of facial expressions. Finally, the control point's vector is used to mimic the artificial face simulator [Prince. M. (2017)].

The parameter is determined as; for the mood previous emotional state has been buffered, for personality Big Five model of personality has been considered [Power R. A. et al. (2015)], for Euclidean intensity distances between the feature vector of standard and user emotion are proposed being used, for speech the training data set of data statistic and machine learning based on Ultra-large-scale database of natural language are used [McGilloway et al. (2000); Greasley (2000); Mohamad Nezami et al. (2019)], and for the colour around eyes RGB color patterns for different emotions have been used [Johnson et al. (2013)].

One of the objectives of most of the research is to improve the life of human beings. So in this regard studying human-machine behaviour becomes essential. Moreover, Human-Computer Interaction (HCI) is one of the areas under which humans and machines should have better communication skills. Ultimately, when we communicate with the humanoid robot, we want to communicate as a companion rather than as a machine closer to human nature.

Here, we are proposing a humanoid robot model that can emotionally respond and a human being. The robot can recognize the user emotional state and would respond accordingly. That could mean that if the user is happy, then the robot should behave like if it is also happy, which would improve the interaction between a human and a machine. The problem arises for the robot to communicate with a user whose expression is not clear on his face, especially when people are involved.

The rest of the paper organized as the related work section is following this section then the complete methodology is presented, the result and discussion section is presented following the methodology and finally the work was concluded.

II. Literature Review

Most research on robotics heads mimicking human facial expression is done in some universities and research institutions of the United States, Japan, and the European Union. A robot called Kismat is one of the examples of it, developed by Cynthia Breazeal at the Massachusetts Institute of Technology. Waseda University, Japan, has developed a series of robots named WE-R since 1996 [Hiroyaus Miwa, (2001)].

In the recent decade, many researchers have been trying to recognise human beings' facial expressions automatically. Various pattern recognition methods have been used in order to recognize facial expressions. As in our previous works [Master Prince, (2013a); Master Prince (2013b)], the novel approach has been introduced to recognize facial expressions. As discussed in [Young. A. W. et al., (1989); Padgett. C. et al. (1997)] reported that approaches to emotional robot design often adopted results from psychology to design robot behaviours to mimic human beings. Miwa et al. proposed a mental model to build the robotic emotional state from external sensory inputs [Miwa. H.et al., (2003), Miwa. H.et al. (2004)]. Duhaut presented a computational model which includes emotion and personality in robotic behaviours [Duhaut. D. (2008)]. Moshkina et al. give a model of time-varying effective response for humanoid robots based on the Traits, Attributes, Moods and Emotions [Moshkina et al. (2011)]. One of the most important aspects is the robot mood transition from the current to the next mood state, which influences the robot's interaction behaviour and a user's feeling. Meng-Ju Han et al. introduced an effective model to make transition among mood states would become smoother and thus might enable a robot to respond with more natural emotional expressions [Meng-Ju Han et al. (2013)]. Imitating emotions with RGB patterns has previously been proposed with other types of robots. For example, Kanoh et al. asked 50 people to identify which of Ekman and Friesen's six basic emotions the Ifbot was imitating with its 29 pre-programmed facial expressions [Ekman P. et al. (1969), Kanoh. M. et al. (2005)]. Angelica Lim and Hiroshi G. Okuno proposed speech and gait analysis to recognize human emotion [Angelica Lim et al. (2012)].

Previous researches have shown powerful tools for designing emotional robots. It is observed that exact emotion recognition and exact mimicking the artificial emotion plays a vital role in effective HRI. These representations lack a theoretical basis to support the assumptions in their emotion recognition design and simulation of the artificial face. That motivates me to investigate an effective emotion recognition system and an excellent artificial face simulator. Emotion recognition system by recognizing the pattern of the facial muscles is not enough. The speech recognition and speech synthesis function modules are embedded as emotion recognition procedures [Jianfeng *et al*. (2019); Mehmet *et al*. (2020)]. The combination of control point vectors and RGB patterns is used to imitate the artificial emotion with a smooth mood transition and get back to its normal intensity state of the current emotion after showing its actual intensity. Questionnaire surveys were conducted to evaluate the effectiveness of the proposed method.

III. Proposed Work

Our previous work proposed a model artificial brain emotion recognition and generation system (ABERGS) [Prince. M. (2017)]. This work is an extension of ABERGS, where facial expression is fused with the voice to simulate artificial expression. In our proposed work, the loud voice recognized by the robot were classified as recognized emotion using the trained CNN model [Jianfeng et al. (2019)] and RGB patterns were used to enhance artificial expression [Johnson et al. (2013)]. The purpose of the RGB pattern is to use it around the robot's eyes to express exact emotion.

In order to find out RGB color patterns against each specific emotion, an experiment was set up.

A. Speech Emotion Recognition

Fig. 1 illustrates the layers which are substituted to the CNN (feature learning block (FLB) and LSTM (Long short-term memory). Four FLBs extract the low-level features of speech, such as emotional features, and LSTM can learn the high-level features, which contain both the local information and the long-term contextual dependencies.

The pretrained model [Jianfeng *et al*. (2019)] was used as transfer learning to get trained with emotional speech dataset [McGilloway *et al.* (2000)], and the accuracy was outstanding, as shown through the confusion matrix in Fig. 2. The idea was to verify the accuracy of the model.

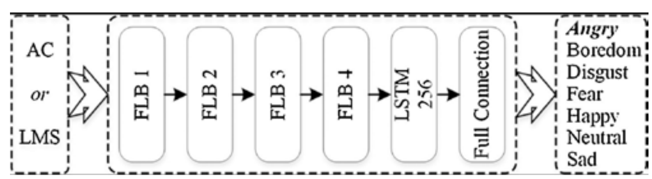


Fig. 1. Block Diagram of the Overall Architecture of the CNN [Jianfeng et al. (2019)].

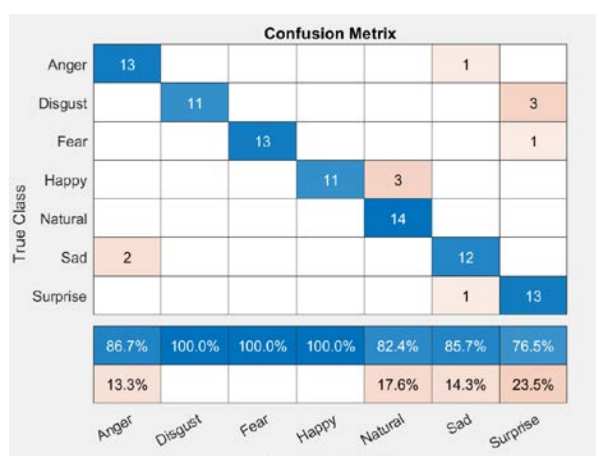


Fig. 2. Confusion Matrix Post-Training Model [Jianfeng *et al*. (2019)] with 98 Samples of [McGilloway *et al.* (2000)] Emotional Speech Dataset.

B. RGB Pattern Recognition

A movie clip against each major emotion (Natural, Happy, Sad, Anger, Surprise, Fear, and Disgust) has been selected. In a small theatre room, 20 participants are invited to watch the movies. A video (pointing to eyes) of each participant has been made against each movie clip. Now we have 7*20 videos (20 videos for each emotion). Now, each video is analyzed as below:

Measurements of the following attributes of each color (R, G and B) for each emotion.

- Intensities: An intensity of emotion.
- Duration: How long does emotion exist?

Intensity calculation: Here, the intensity is measured in terms of the RGB scheme. Each frame of the video is converted into a grayscale color scheme. Each color's intensities (R, G and B) can range between 0-255.

$$^{H}I_r = \sum_{f=1}^{20} r/20 \tag{1}$$

$$^{H}I_g = \sum_{f=1}^{20} g/20 \tag{2}$$

$$^{H}I_b = \sum_{f=1}^{20} b/20 \tag{3}$$

The above equations show the average intensity of each color pattern (R, G and B) against happy emotion. Here R, G and B value ranges from 0 to 255. The same intensities of each color pattern (R, G and B) for all other emotions can be calculated.

Duration calculation: The time duration between emotions initiated and coming back to the normal state. The RGB value for normal emotion has got 150, 150, and 0, respectively. It consists of two things;

- Rise time: Time is taken to rise from normal to peak value.
- Fall time: Time is taken to return to the normal from the peak value.

$$D_H = D_r + D_f \tag{4}$$

where D_H is the total duration of an emotion existing on the face, D_r is the time taken to rise from normal to peak, and D_f is the time taken to return to the normal from the peak value. Same as intensity, duration is also calculated on average.

After analysis, all the videos, follow Table I is obtained. Which shows duration, color (RGB), rise and fall time for all recognized emotions. Table I: RGB color pattern's Intensities, Duration, Rise time and Fall time.

C. RGB Pattern Evaluation

Table I from the previous experiment determined which pattern of color intensity and duration of the RGBs that humans associate with specific emotions. The purpose of this experiment is to examine whether humans recognize these RGB patterns as emotions or not. Two RGB patterns for each emotion are based on Table I and glow through the computer screen, as shown in Fig. 3.

TABLE I. RGB COLOR PATTERN'S INTENSITIES, DURATION, RISE TIME AND FALL TIME

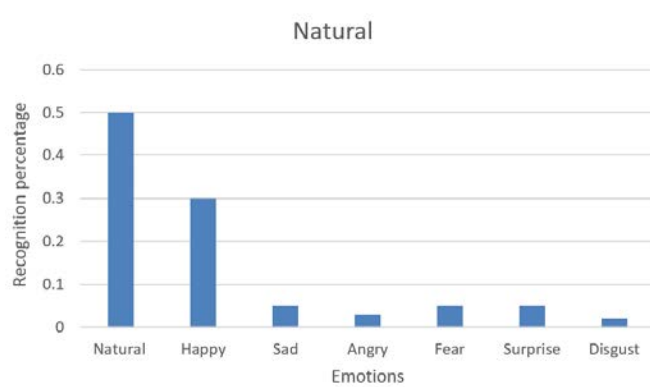
Emotion	Duration (Sec)	Color(RGB)	Rise time (%)	Fall time (%)
Normal	0	150, 150, 0	0	0
Anger	1.7	255, 0, 0	83	17
Surprise	4.0	255, 255, 0	83	17
Disgust	2.3	0, 75, 0	83	17
Sadness	4.3	0, 100, 225	20	80
Happiness	2.1	200, 120, 0	80	20
Fear	4.0	0, 20, 80	20	80

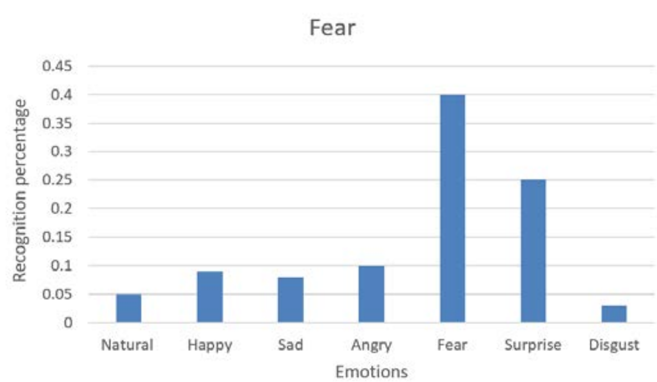
Fig. 3. RGB Patterns shown on the Computer Screen to Recognize Emotion for RGB Pattern Evaluation.

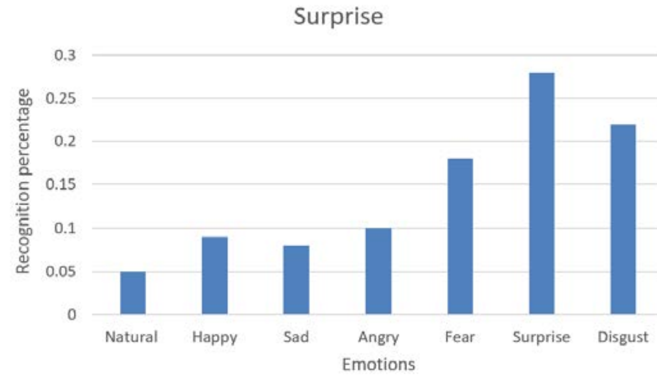
For this, fifty participants were invited, and a survey was conducted. They were asked to check on the recognized emotion after seeing the RGB pattern on the computer screen into robot eyes, and the patterns were shown in random order to avoid ambiguity.

The responses of the participants are interpreted into a bar chart, as shown in Fig. 4. The result was significant as all the emotions were recognized. Moreover, some ambiguity was witnessed between Natural and Happy, Sad and Angry, and Fear and Surprise.

Finally, the data has been collected and analyzed.







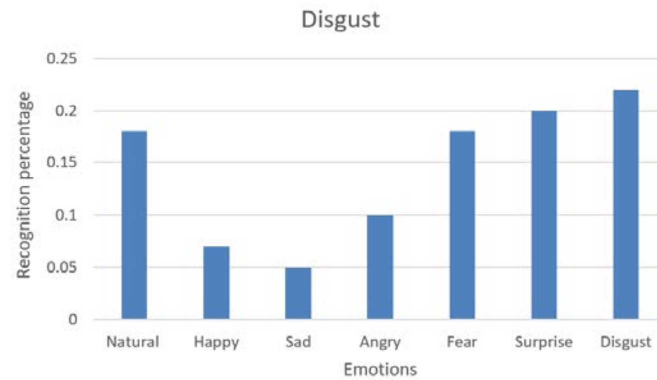


Fig. 4. Average Degree of Recognition for RGB Patterns against each Emotion.

IV. Result and Discussion

The result shows a moderate degree of recognition against each emotion. As a result, it was a bit ambiguous, but when it combined with ABERGS [Prince. M. (2017)], the result was very effective. Fig. 5 illustrates the simulated image.

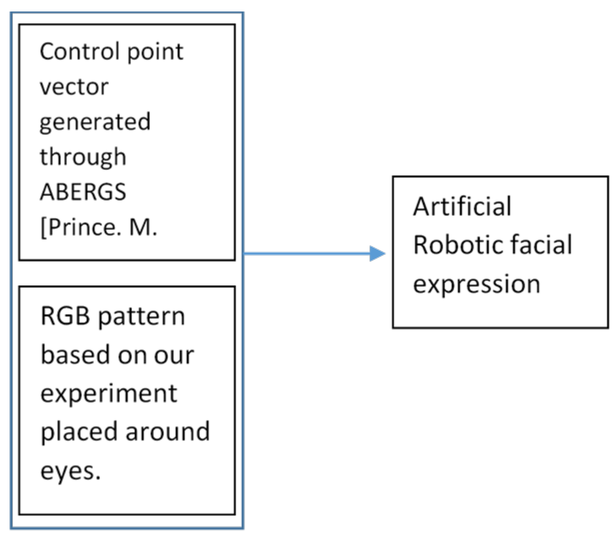


Fig. 5. ABERGS [Prince. M. (2017)] and the RGB Pattern used simultaneously to Mimic Robotic Facial Expression.

TABLE II. COMPARATIVE SATISFACTION LEVEL OF ABERGS [PRINCE. M. (2017)] VS ABERGS [PRINCE. M. (2017)] WITH RGB

Emotions	**Satisfaction Level (without voice)**	**Satisfaction Level (With voice fusion)**	**Satisfaction Level (With RGB Pattern)**
Natural	96.5%	98%	98.7%
Happiness	97%	98%	99%
Surprise	96.5%	97.5%	98%
Fear	97.5%	99%	99.2%
Sadness	96%	98%	98.3%
Disgust	97.5%	98.5%	99%
Anger	95.0%	98%	99.8%

Table II illustrates the effectiveness of the RGB pattern with ABERGS [15].

V. CONCLUSION AND FUTURE WORK

The proposed model uses facial image, voice and simulate artificial emotion on a robotic face with an RGB pattern in the eyes. With the first experiment and RGB patterns were determined, and with the second experiment, the model's effectiveness was tested. The results were very effective. For future work, human gait can also be considered in order to recognize the emotion of the human being. Second, we will focus more on processing speed by using GPU with an efficient algorithm.

ACKNOWLEDGMENT

The author would like to gratefully acknowledge Qassim University, represented by the Deanship of Scientific Research, for the support for this research under the number 1319-coc-2016-1-12-S.

REFERENCES

[1] Angelica Lim and Hiroshi G. Okuno, (2012): Using speech data to recognize emotion in human gait", A. A. Salah et al. (Eds.): HBU, LNCS 7559, pp. 52-64, Springer-Verlag Berlin Heidelberg.

[2] Duhaut D. (2008): A generic architecture of emotion and personality, in Proc. IEEE INT. Conf. Adv. Intell. Machatron., Xian, China, pp. 188-193.

[3] Ekman P. Friesen WV (1969) The repertoire of nonverbal behaviour: categories, origins, usage, and coding. Semiotics 1:49-98.

[4] Greasley P, Sherrard C, Waterman M (2000) Emotion in language and speech: Methodological issues in naturalistic approaches. Lang Speech 43:355–375.

[5] Hiroyaus Miwa, Tomohiko Umetsu, Atsuo Takanishi, et al. (2001): Human2 like robot head that has olfactory sensation and facial color expression" Proceeding of the 2001 IEEE, ICRA, and @001: 459-464.

[6] Jianfeng Zhao, Xia Mao, Lijiang Chen (2019): Speech emotion recognition using deep 1D & 2D CNN LSTM networks, Biomedical Signal Processing and Control, Volume 47, Pages 312-323, ISSN 1746-8094.

[7] Johnson, D.O., Cuijpers, R.H. & van der Pol, D. (2013). Imitating Human Emotions with Artificial Facial Expressions. Int J of Soc Robotics 5, 503–513.

[8] Kanoh M, Iwata S, Kato S., Itoh H (2005) Emotive facial expressions of sensitivity communication robot "Ifbot". Kansei Eng Int 5(3):35-42.

[9] Master Prince, (2013a): A Simple Method for Face Normalization Based on Novel Normal Facial Diagram", International Journal of Video & Image Processing and Network Security IJVIPNS-IJENS Vol:13 No:03.

[10] Master Prince, (2013b): Enhancing face normalization based on novel normal facial diagram, The 26th International Conf. On Computer Applications in Industry and Engineering (CAINE-2013), Los Angeles, California, USA, pp. 97-100, 25-27 September 2013.

[11] McGilloway S, Cowie R, Douglas-Cowie E, Gielen S, Westerdijk M, and Stroeve S (2000) Approaching automatic recognition of emotion from voice: A rough benchmark. In: Proceedings of ISCA workshop on speech and emotion, Belfast, United Kingdom.

[12] Mehmet Berkehan Akçay, Kaya Oğuz, Speech emotion recognition (2020): Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers, Speech Communication, Volume 116, Pages 56-76, ISSN 0167-6393.

[13] Meng-Ju Han, Chia-How Lin, and Kai-Tai Song, (2013): Robotics emotional expression generation based on mood transition and personality model, IEEE Transactions on Cybernetics, VOL. 43, NO. 4, August.

[14] Miwa. H., Itoh. K., Matsumoto. M., Zecca M., H. takanobu, S. Rocella, M. C. Carrozza, P. Dario, and A. Takanishi, (2004): Effective emotional expressions with expression humanoid robot WE-4RII: Integration of Humanoid Robor hand RCH-!," in Proc. IEEE/RSJ Int. Conf. Intell. Robots. Syst., Sendai, Japan, pp. 2203-2208.

[15] Miwa. H., Okuchi. T., Itoh. K., Takanobu. H., and Takanishi. A. (2003): A new mantal model for humanoid robots for human friendly communication introduction of learning system, mood vector and second order equations of emotion, in Proc. IEEE Int. Conf. Robots Autom., Taipei, Taiwan, pp3588-3593.

[16] Mohamad Nezami, O., Jamshid Lou, P. & Karami, M. ShEMO, (2019).: a large-scale validated database for Persian speech emotion detection. Lang Resources & Evaluation 53, 1–16.

[17] Moshkina. L., Park. S. , Arkin. R. C., Lee J. K., and JungH. (2011): TAME: Time varying affective response for humanoid robots, Int. J. Social Robot., vol. 3 no. 3. 207-221.

[18] Padgett. C. and Cottrell. G. (1997): Representing face images for classification emotions, vol-9. Cambridge, MA: MIT press.

[19] Power RA, Pluess M. (2015): Heritability estimates of the Big Five personality traits based on common genetic variants. Transl Psychiatry.;5:e604.

[20] Prince. M. (2017): Adaptive artificial brain for humanoid robot using Pattern Recognition and Machine learning, IJCSNS International Journal of Computer Science and Network Security, VOL.17 No.5, May.

[21] Yang, Y., Ge, S.S., Lee, T.H. et al. (2008), Facial expression recognition and tracking for intelligent human-robot interaction. Intel Serv Robotics 1, 143–157.

[22] Young. A. W and Ellis. H. D. (Editors) (1989): Handbook of Research on Face Processing. North-Holland, Amsterdam: Elsevier Science publishers B. V., ISBN 0 444 87143 8.

AUTHORS' PROFILE

Master Prince, received the B.S degree in computer science from Patna University, India, in 1996, the M.S degree in computer science from Indira Gandhi National Open University, New Delhi, India, in 2004, and the Ph.D. degree in computer science from Pune University, India, in 2008.

Since 2009, he has been working as an Assistant Professor with the Department of Computer Science, Qassim University, Saudi Arabia. His research interests include computer vision and machine learning.

Dr. Prince received the Best Ph.D. Thesis Dissertation of the Year 2009 Award of the Pune University, India.