Workshops at 18th International Conference on Intelligent Environments (IE2022)
H.H. Alvarez Valera and M. Luštrek (Eds.)

doi:10.3233/AISE220020

Multimodal Emotional Understanding in Robotics

Juanpablo HEREDIA [a,1], Yudith CARDINALE [b,c], Irvin DONGO [b,d], Ana AGUILERA [e], Jose DIAZ-AMADO [b,f]

[a] *Computer Science Department, Universidad Católica San Pablo, Arequipa, Peru*
[b] *Electrical and Electronics Engineering Department, Universidad Católica San Pablo, Arequipa, Peru*
[c] *Universidad Internacional de Valencia, Spain*
[d] *Univ. Bordeaux, ESTIA INSTITUTE OF TECHNOLOGY, Bidart, France*
[e] *Escuela de Ingeniería Informática, Facultad de Ingeniería, Universidad de Valparaíso, Valparaíso, Chile*
[f] *Electrical Engineering, Instituto Federal da Bahia, Vitoria da Conquista, Brazil*

Abstract. In the context of Human-Robot Interaction (HRI), emotional understanding is becoming more popular because it turns robots more humanized and user-friendly. Giving a robot the ability to recognize emotions has several difficulties due to the limits of the robots' hardware and the real-world environments in which it works. In this sense, an out-of-robot approach and a multimodal approach can be the solution. This paper presents the implementation of a previous proposed multimodal emotional system in the context of social robotics; that works on a server and bases its prediction in four modalities as inputs (face, posture, body, and context features) captured through the robot's sensors; the predicted emotion triggers some robot behavior changes. Working on a server allows overcoming the robot's hardware limitations but gaining some delay in the communication. Working with several modalities allows facing complex real-world scenarios strongly and adaptively. This research is focused on analyzing, explaining, and arguing the usability and viability of an out-of-robot and multimodal approach for emotional robots. Functionality tests were applied with the expected results, demonstrating that the entire proposed system takes around two seconds; delay justified on the deep learning models used, which are improvable. Regarding the HRI evaluations, a brief discussion about the remaining assessments is presented, explaining how difficult it can be a well-done evaluation of this work. The demonstration of the system functionality can be seen at https://youtu.be/MYYfazSa2NO.

Keywords. Emotional understanding, Human-robot interaction, Multimodal method

1. Introduction

Emotional understanding is a challenging task in daily human communication and Human-Robot Interaction (HRI) [1,2]. Several data sources help determine an emotion concerning something, such as products, people, or the interaction itself. In response to the multiple data analyzable, multimodal machine learning and deep learning methods have achieved high performance in robotics and other areas.

[1] Corresponding Author; E-mail:juanpablo.heredia@ucsp.edu.pe

A multimodal approach aims to take advantage of each available data source [2,3]. It could be a complex method, although this has been simplified to be applicable in small or straightforward environments, such as robots. Common data types used are visual and auditory, for instance, through videos of people speaking. However, thanks to the robots' sensors, other types of data might be used [2], such as thermal facial images, with valuable data of face muscles; body pose and kinematics that could capture the body language; voice, which allows analyzing both how and what people say something; or brain activity, with very revealing information about the emotional state of people, although it requires more specific and expensive sensors.

Emotional understanding of robots does not include only people's emotional recognition but also the capability to infer people's emotional state or even have and express an artificial emotional state [2]. In this sense, this research present the implementation of a multimodal emotion recognition method [4] in the context of social robotics. The complete implementation can be understood as a system that involves the input data captured with the robot's sensors, the emotion recognition process, and the triggering of simple changes to the robot's behavior, which would adapt accordingly. The employed modalities are four: face expressions, posture, body features, and context features.

Moreover, a robotic emotional system must be evaluated; however, the HRI experiments can become complex, requiring several experiment subjects (users), established questionnaires or interviews. Functionality tests were carried out inside the laboratory with the expected results. These experiments demonstrated that the system could work in around two seconds from the image input capture to the robot's behavior changes, where the deep learning processing consumes more than a second and a half. However, the deep learning-based emotions recognition method is upgradeable. Because the system is emotional, knowing the actual experiment subjects' emotions and feelings is necessary, making experiments even more difficult. The scope of the current research work does not cover the experiments but discusses the possible ones; some of them demand real scenarios where people express real emotions (not faked or imitated). Therefore, the contribution of this work is two fold: first an initial implementation of an end-to-end emotional robotic system, and second a slight clarification of how to evaluate the HRI systems. Besides, this paper seeks to continue the research towards emotional understanding between robots and humans, thus allowing the analysis of that interaction.

The organization of this work is followed by the related work of emotion recognition in robotics and the multimodal approach, described in Section 2. Then, the explanation of the proposed emotional system for robots is presented in Section 3. Section 4 explains the details of the environment and implementation and technical results in terms of execution time. The discussion about evaluation and experiments is exposed in Section 5. Finally, the conclusions reached are presented in Section 6.

2. Related Work

The most common technique for emotion recognition from multimedia data is by neural networks, specifically the convolutional and the recurrent ones. In robotics, the models used must be simple or optimized due to the limited hardware. For instance, Ghoshal et al. [5] propose a face recognition and emotion detector based on short deep learning architectures and functionalities of OpenCV for a pet robot. Although this proposal works in real-time, it is limited by data conditions because of the training data for facial emo-

tion detection. This model could not work in real-world applications where faces are not appropriately captured. Huang et al. [6] consider more modalities and use posture and context information to realize emotions in videos. They propose a three-branch architecture with three feature streams: body images, processed by a 3D-ResNet101; skeleton graphs, processed by the Actional-Structural Graph Convolution; and context images (where the person target is removed from the whole image), processed by a convolutional network of five blocks. This method does not perform in real-time but achieves effective emotion recognition in public space video data; this kind of method described in [5,6] has a complicated implementation due to the complex architecture. Therefore, topics such as transfer learning have helped, as shown by Webb et al. [7] using a convolutional neural network pre-trained as a stacked convolutional autoencoder that achieves a good performance in unconstrained environments. This model achieves state-of-the-art results in recognizing facial emotions and attending face images with lighting and pose variations. Thus, a transfer learning-based model can be suitable for real-world applications, showing their results with data collected by a Nao robot.

Working with the robot as just the inputs catcher to do the processing outside the robot (e.g., in a server in the same network or different device on the robot) allows the execution of more complex methods for detecting emotions. Moreover, it makes the treatment of multimodal data more comfortable because of the gain of resources for processing. Greco et al. [8], for example, add a device (Nvidia Jetson TX2) with a graphics card into the Pepper robot that performs face detection and emotion recognition, and the robot modifies its voice and movements. The architectures tested are MobileNet and ResNet-18 with a temporal module; in the end, their better results are achieved by a ResNet-18 plus a temporal average module. They argue that the robot can perform a good emotion recognition process in real-time by adding an extra device. Unlike Greco et al. [8], Martinez et al. [9] use an external server instead of the extra device. The models used might be more complex, and their updating or deployment would require more effort. Martinez et al. [9] use drones for capturing the inputs but do not send feedback but analyze and adapt future responses to different situations. For emotional recognition, they use a convolutional neural network. Their method is evaluated in a simulated environment, but the processes have been tested with data from the real world; thus, the deployment in a real scenario should work similarly.

Regarding multimodal emotion recognition methods for robotics, it is becoming more popular and has better results. Heredia et al. [10] present a method thought and applicable to robots that use images and speech data to predict an emotion. Although this model receives two types of data (image and audio), it processes three modalities because the audio is transcribed and works as a text modality. Liu et al. [11] present a multimodal system that employs different inputs such as speech, facial expression, body gesture, eye movement, and psychological signals. Each data modality is processed individually until getting independent predictions, and then a Bayesian classifier is used to get a multimodal result. They implement their system with several robots communicating with a server (which performs all emotion recognition processes and sends feedback). The evaluation of the system is carried out in four scenarios, including tourist guides, entertainment games, family services, and simulated scenarios, which demonstrate a fluid interaction between humans and robots. Liu et al. [11] seek to recognize emotions and endow emotional communication ability for robots by showing emotional behaviors with images, voice, movements, and other available ways. Yu and Tapus [12] propose an inter-

active robot learning framework. This method uses multimodal data from thermal facial images and human gait data for online emotion recognition and a fusion method for the multimodal classification using the Random Forest model.

Similarly, Bera et al. [13] use different data inputs captured from different perspectives, from the robot's camera and a surveillance camera. They propose a system for socially-aware navigation to analyze the emotional behavior from people's faces and trajectories. In this work [13], the robots do not perform behavior changes because of emotion detected, but they support and adjust their navigation system according to a detected feeling.

Moreover, the robot's multimodal emotion understanding can also be applied to robots, changing several features according to the emotions detected. Loffler et al. [14] make robots express artificial emotions in response to a people's feelings recognized using color, motion, and sound. This study is focused on the user interaction and perception with the robots; 33 persons participated in experiments where they passed a mood questionnaire after interacting with a robot. The subjects chose what emotion they thought the robot expressed and rated the confidence. The results show that using a multimodal communication of emotions is better in HRI; in addition, the combination of color and movement is better for the communication in terms of cost/benefits ratio.

This research presents a more exhaustive emotional image processing by analyzing four modalities from only one image and a more complex fusion method than the reviewed works because of its novelties. The feasibility of out-of-robot approaches and the multimodal methods in the robotics context is also shown in this work. Using just images, up to four types of inputs (modalities) can be obtained, adding more modalities without significantly affecting performance because almost all processing is located in a server. In addition, the emotional response of the robot is multimodal and responds to previous research on HRI.

3. Multimodal Emotion Recognition System

There are two main options for developing an emotional system for robots: inside the robot implementation or outside the robot implementation. Both options have their advantages, summarized in terms of response speed, method's complexity, and accuracy of results.

In an implementation inside robots, the response time for recognizing emotions and making decisions over the results might be short. It depends on the available hardware that limits the machine learning methods used. Moreover, the simpler models used, the lower accuracy of results is expected. However, the final results could be improved using several modalities (e.g., faces, body posture, gestures, context from images and videos, voice, text) [4,10,11].

On the other hand, an implementation outside robots has the response time as a weakness because it depends on the connection between the robot and the external device (server). Regarding the machine learning methods used, using a server, the hardware is not a problem because it is easier to improve; also, the achieved accuracy is expected to be better than inside approaches. Another advantage of the outside approaches is the ability to be implemented with different robots or even implement complex systems with several robots simultaneously. Besides, developing the emotion system outside leaves more resources for fundamental robot applications, such as mapping, navigation, and

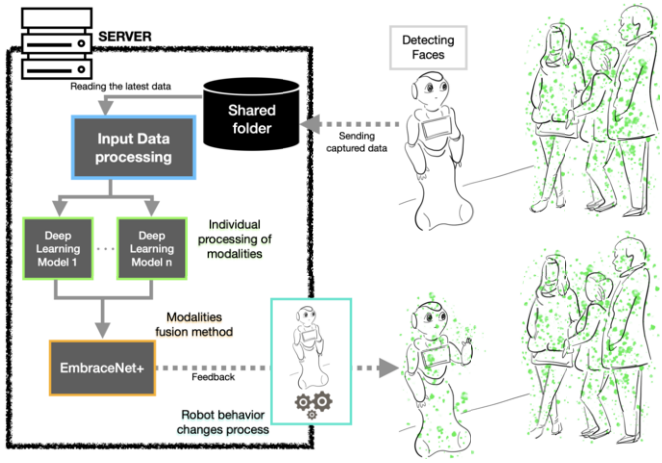

Figure 1. Overall pipeline, pointing out the components of the implemented system.

routing tasks. With hybrid approaches, as presented in [8], the addition of extra devices inside the robot does not have the flexibility in the deployment environment.

In this work, it is implemented a multimodal emotion recognition system [4], following the approach outside the robot, which can use the recognized emotion to adapt its behavior. The general pipeline of the system is summarized in Figure 1. It is roughly composed of four components: the input data processing, the individual processing of modalities, the modalities fusion method, and the robot behavior change process. These components are detailed in the following subsections separately.

3.1. Multimodal Emotion Recognition Method

This system's component comprises three parts, input data capturing and processing, the processing of individual modalities, and the fusion of modality results. In our current version of the system, we have implemented the method proposed in [4], which uses modalities extractable from images, such as facial images, body images, context images, and body postures.

3.1.1. Input Data Processing

This part is vital because it defines the valuable data that will allow emotional understanding. Almost every modality needs preprocessing for being the input of the method. In some cases, these procedures might be machine learning methods as complex as the emotion recognition ones. Heredia et al. [4] use three deep learning models to get the proper inputs (YOLOv3, RetinaFace, and High-resolution neural network). In this part, a data discriminator could also be used to detect non-applicable, unavailable, or low-quality data; for example, if the faces are illegible or the recorded speeches have much noise, this could avoid errors and waste of resources. Regarding the functionality, the robot captures the raw data with its sensors and sends it to a shared folder to be read and processed by the server; hence, the server constantly reviews and takes the most recent data storage. The robot is programmed to capture and send images only when the robot detects a face to avoid unnecessary executions.

3.1.2. Individual Processing of Modalities

Inside the server, deep learning models process each modality's data. In [4], to get individual predictions, there are used three types of networks. A VGG-face model processes the face images, and the postures abstracted into graphs are processed by the direct graph convolutional neural network. Then, the body and context modalities are processed separately by the attention branch network, that weights the image's areas with greater emotional meaning without considering the target person. This part aims to extract features from each input data while obtaining preliminary predictions. Thus, in [4], each modality outputs two vectors, a prediction vector of n values (where n is the number of classes of emotions) and a vector of intermediate data with 512 values.

3.1.3. Modalities Fusion Method

To merge the individual results, there exist several approaches. A simple approach could be an average between individual predictions, but nowadays, a machine learning-based method achieves betters performance, especially for real applications. In the method proposed in [4], there is used the technique EmbraceNet+, which inherits the advantages of the basic EmbraceNet [15] and improves some aspects. Both are based on the multinomial distribution for merging feature vectors into a single one, and some linear layers are used for learning the correlation between modalities. The big difference is that EmbraceNet+ has a more complex architecture; it comprises three embracenet models and two additional simple fusion methods (concatenation and weighted sum). Moreover, thanks to using an availability values vector as a parameter, modalities' adaptability, and the filtering of valuable data are achieved.

3.2. Robot Behavior Changes Process

How the recognized emotion is used for decision making, and behavior adaptation of robots is perhaps the most critical part of an emotional system because it shows the results of HRI. If this component is complex and well-done enough, the robot could express emotions to show empathy. This part should be implemented inside the robot and work after receiving the feedback of emotions detected from the server.

For the presented emotional system, simple behavior changes are implemented to show empathy by imitating the recognizing emotions. The results of the used multimodal emotions recognition method [4] are vectors of eight boolean values, respective to anger, anticipation, disgust, fear, joy, sadness, surprise, and trust; moreover, there could be more than one emotion detected, and there could be as vectors as the number of persons in images. Thus, the decisions making process is set as a voting model where positive emotions increase the *emotion value*, and negative emotions will decrease it. Also, some emotions are stronger than others, so the value for increasing or decreasing can differ. The scheme used in our current version is: positive emotions joy (+2), trust (+2), and anticipation (+1), surprise (+1); negative emotions sadness (-1), fear (-1), anger (-2), and disgust (-2). For example, if joy, surprise and anger are detected, the *emotion value* would be 1 ($+2+1-2$). Finally, two thresholds are set in +3 and -2, allowing the robot to define three emotional states. These states are: (i) Negative, when the *emotion value* is less or equal to -2; (ii) Neutral, when the *emotion value* is more than -2 and less than +3; and (iii) Positive, when the *emotion value* is more or equal than +3.

Both the values of each emotion and the defined thresholds are configurable. They can be established differently following the possible experiments and the context and tasks performed by robots.

4. Implementation Details and Functional Results

The implementation of the proposed system was tested with a Pepper robot. This human-like robot has a head with microphones × 4, RGB cameras × 2, one 3D sensor, and touch sensors × 3; a chest with a gyro sensor; two arms and hands with a touch sensor in each one; legs (which is one) with sonar sensors × 2, laser sensor × 6, bumper sensors × 3, and one gyro sensor. The Pepper robot has 20 motors to move its head, shoulders, elbows, wrist, hands with five fingers, hip, knee, and base. Besides, this robot has a Wi-Fi connection with frequency bands of 2.4GHz to 5GHz.

The server, where the system is running, has two video cards Nvidia RTX-2080, a processor intel core i9-9820X, 64Gb of RAM; moreover, it is installed Ubuntu 18.04. The software environment uses Python, Python2.7.16 and Python3.9.8 since the NAOqi API is used and works only in Python2, but all used deep learning models need Python3 with PyTorch version 1.8.2. The NAOqi API allows communication with Pepper robots asking for services. Thus, getting the input images requires two modules from NAOqi, ALFaceDetection and ALVideoDevice. ALFaceDetection is used to program the robot for detecting people's faces. When it occurs, it can be used ALVideoDevice to access the robot's camera and get the input image. Similarly, once the emotion is detected and behavior changes are determined, in the current version of our model, just two actions are taken: move the robot forward or backward and show an image in an annexed display device. Thus, the modules ALMotion (for moving the robot) and ALTabletService (for showing images on the tablet's screen) are used to make the robot execute the changes.

The chosen colors of images displayed on the tablet respond to a color study where the meaning of colors is qualitatively explored [16]. In this study, the yellow color is more robust and can be cataloged as smiley; the blue (and green) are more soft colors and can relax people by reminding the sky or forest; finally, gray, that is a neutral, weak or even a dull color. Thus, yellow is designated for the Positive state, blue for the Neutral one, and Gray for the Negative state. Other and more complex changes or actions can be executed for the robot to be more emotional, according to its capacities. For instance, the robot can move its arms, hands, and head, change its eyes' color, and modulate its voice. However, these changes must be in harmony with the context and the task to perform. For instance, eccentric movements of arms and a funny voice can seem out of place for a museum guide robot.

The three possible states were tested in terms of functionality with successful performance. However, since the context and body properties were the same in experiments, the analyzed face expressions and posture had to be clear and obvious (with-

Figure 2. Reaction from the Neutral scenario. The green frame show the robot behavior, and the pink frame the actual picture capture by robot's camera.

Figure 3. Reaction from the Positive scenario (a) and Negative scenario (b). The green frame show the robot behavior, and the pink frame the actual picture capture by robot's camera.

out being exaggerated) to achieve the desired functioning. Figure 2 shows the robot's behavior in the Neutral state, which is defined as a slow movement forward and the showing of a blue image with a neutral face emoticon; Figure 3(a) shows the robot's behavior in the Positive state, which is defined as a normal movement forward and the showing of a yellow image with a happy face emoticon; and Figure 3(b) shows the robot's behavior in the Negative state, where is defined a slow movement backwards and the showing of a gray image with a serious face emoticon. In another iteration of tests (https://youtu.be/MYYfazSa2N0), some flaws were noted that might be refined in the future. Due to the emotion recognition method analyzes the whole picture, not just the centered person's face, if there are several persons the accumulated *emotional value* tends to be high (positive), thus the *emotional value* calculation could improve and become sophisticated. In addition, the facial emotion detection deep learning model has weaknesses if the person is wearing glasses or masks.

In terms of time spent, receiving input images takes ∼0.24 seconds, while the communication for ordering robot behavior changes takes ∼0.06 seconds. The multimodal emotion recognition method (including the input processing, the processing of individual modalities, and the fusion of modality results) takes, on average, ∼1.9 seconds. However, in the first time execution, because of the load of deep learning models, it takes ∼4.6 seconds. The whole system takes approximately 2.22 seconds. Although the communication time between robot and server allows a real-time (or near real-time) data exchange, the whole processing of images takes more time, making it difficult to act in real-time and could hinder the HRI. However, improving and updating deep learning models can improve system performance, in terms of execution time and results in complex scenarios.

5. Discussion on Evaluation an Emotional Robotic System

Evaluating the robot's performance for recognizing emotions is counterproductive because the models used are external, and they can be the best possible model. Thus, the evaluations must focus on the complete emotional system in subjective terms, for example, people's comfort when interacting with the robot. The importance of these evaluations has been considered a challenge in HRI [17]. To carry them out, a specific environment and task must be established following the acceptability and usability of the end-users, to be exploited in real applications [17].

The task performed by users and robots can be simple, such as small talks between people and robots, or more complex, such as robot museum guides that explain and answer people's doubts. Thus, the system should be tested in natural settings considering

the very complex emotions of humans, including aspects such as socio-cultural background [17]. In addition, to know the real emotions of people, surveys or interviews may not be enough, and some biological sensors could help by measuring temperature, heart rate, or even brain activity [18].

Within the aspects that can be measured to demonstrate the excellent behavior of the robot are: the Task Effectiveness, which assesses how well a human-robot team accomplishes established tasks; The Robot Attention Demand, metric that can be calculated with the relation between the Neglect Tolerance value of the robot and the Interaction Effort by the person [19]. Thus, HRI could be improved by increasing carelessness tolerance and decreasing interaction effort. Although each metric said can be applied separately, together, they provide an HRI assessment framework [19].

Other experimentation could be a sophisticated simulation [9]. In this experiment, worse and weird cases can be tested without affecting real persons; however, the results will be conditioned without much rigor. A simulation can be a helpful tool for developing but not enough for evaluating an HRI system.

Ultimately, this research work also nominates complex experimentation intending to analyze how people feel when interacting with the robot and how much a robot can influence and change people's feelings. Additional robot actions may be used to show empathy, such as changing eye color, arm movements, and voice changes. It is planned to program the robot and integrate knowledge into a chatbot to be able to talk to people and provide explanations (for example, a short description of a painting). In addition, it is planned to work with several testers (around 15 persons) with demographic differences and the help of some emotional experts that allow a correct emotional evaluation. These experiments could be carried out in the context of a museum.

6. Conclusions

This work presents an emotional system for robots, developed on a server with connections with robots to get input data and send feedback on the recognized emotions. The system comprises three parts, the capturing and processing of input data, the multimodal emotion recognition method [4], and the robot behavior changes to process.

A well-done emotional system can significantly impact the HRI, giving robots humanized faculties. This work contributes to the progress of the emotional understanding in robots, specifically under the multimodal approaches for capturing real-world data and interacting with humans. In future works, a complete and suitable evaluation will take place. For this, it is considering working with subjects with demographic differences and experts in the area of emotions. Besides, the tentative scenario to perform the experiments might be a museum where the robot can talk or guide people through it and change its behavior according to the detected people's emotions.

Acknowledgement

This research was supported by FONDO NACIONAL DE DESARROLLO CIENTFICO, TECNOLGICO Y DE INNOVACIN TECNOLGICA - FONDECYT as executing entity of CONCYTEC under grant agreement no. 01-2019-FONDECYT-BM-INC.INV in the project RUTAS: Robots for Urban Tourism centers, Autonomous and Semantic-based.

References

[1] Mohammed SN, Hassan AKA. A Survey on Emotion Recognition for Human Robot Interaction. Journal of computing and information technology. 2020;28(2):125-46.

[2] Spezialetti M, Placidi G, Rossi S. Emotion recognition for human-robot interaction: recent advances and future perspectives. Frontiers in Robotics and AI. 2020;7.

[3] Baltrušaitis T, Ahuja C, Morency LP. Multimodal machine learning: A survey and taxonomy. IEEE transactions on pattern analysis and machine intelligence. 2018;41(2):423-43.

[4] Heredia J, Cardinale Y, Dongo I, Díaz-Amado J. A Multi-modal Visual Emotion Recognition Method to Instantiate an Ontology. In: 16th International Conference on Software Technologies. SCITEPRESS-Science and Technology Publications; 2021. p. 453-64.

[5] Ghoshal AM, Aspat A, Lemos E. OpenCV Image Processing for AI Pet Robot. International Journal of Applied Sciences and Smart Technologies. 2021;3(1):65-82.

[6] Huang Y, Wen H, Qing L, Jin R, Xiao L. Emotion Recognition Based on Body and Context Fusion in the Wild. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 3609-17.

[7] Webb N, Ruiz-Garcia A, Elshaw M, Palade V. Emotion Recognition from Face Images in an Unconstrained Environment for usage on Social Robots. In: 2020 International Joint Conference on Neural Networks (IJCNN). IEEE; 2020. p. 1-8.

[8] Greco A, Roberto A, Saggese A, Vento M, Vigilante V. Emotion analysis from faces for social robotics. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE; 2019. p. 358-64.

[9] Martínez A, Belmonte LM, García AS, Fernández-Caballero A, Morales R. Facial Emotion Recognition from an Unmanned Flying Social Robot for Home Care of Dependent People. Electronics. 2021;10(7):868.

[10] Heredia J, Lopes-Silva E, Cardinale Y, Diaz-Amado J, Dongo I, Graterol W, et al. Adaptive Multimodal Emotion Detection Architecture for Social Robots. IEEE Access. 2022;10:20727-44.

[11] Liu ZT, Pan FF, Wu M, Cao WH, Chen LF, Xu JP, et al. A multimodal emotional communication based humans-robots interaction system. In: 2016 35th Chinese Control Conference (CCC). IEEE; 2016. p. 6363-8.

[12] Yu C, Tapus A. Interactive robot learning for multimodal emotion recognition. In: International Conference on Social Robotics. Springer; 2019. p. 633-42.

[13] Bera A, Randhavane T, Prinja R, Kapsaskis K, Wang A, Gray K, et al. How are you feeling? multimodal emotion learning for socially-assistive robot navigation. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). IEEE; 2020. p. 644-51.

[14] Löffler D, Schmidt N, Tscharn R. Multimodal expression of artificial emotion in social robots using color, motion and sound. In: 2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE; 2018. p. 334-43.

[15] Choi JH, Lee JS. EmbraceNet: A robust deep learning architecture for multimodal classification. Information Fusion. 2019;51:259-70.

[16] Clarke T, Costall A. The emotional connotations of color: A qualitative investigation. Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur. 2008;33(5):406-10.

[17] Cavallo F, Semeraro F, Fiorini L, Magyar G, Sinčák P, Dario P. Emotion modelling for social robotics applications: a review. Journal of Bionic Engineering. 2018;15(2):185-203.

[18] Young JE, Sung J, Voida A, Sharlin E, Igarashi T, Christensen HI, et al. Evaluating human-robot interaction. International Journal of Social Robotics. 2011;3(1):53-67.

[19] Olsen DR, Goodrich MA. Metrics for evaluating human-robot interactions. In: Proceedings of PERMIS. vol. 2003; 2003. p. 4.