Social robots: an open-source framework for personal assistant robots

Mohammad Javad Ranjbar Kalahroodi
School of Electrical and Computer Engineering
University of Tehran
Tehran, Iran
mohammadjranjbar@ut.ac.ir

Mohammad Bagher Menhaj
Department of Electrical Engineering
Amirkabir University of Technology
Tehran, Iran
menhaj@aut.ac.ir

Hassan Taheri
Department of Electrical Engineering
Amirkabir University of Technology
Tehran, Iran
htaheri@aut.ac.ir

***Abstract*—From the earliest days of introducing robots, social robots stood as one of the most popular subjects in the human-robot interaction field. These robots are considered to be the future companions of humans and have been used in various fields from entertainment to health care, and it is conceivable that their presence will increase gradually. Despite recent technological advances, these robots still require much development in software and hardware. Social robots should be cost-effective for widespread application. The interaction experience should be engaging in a way to attract users to continue their utilization. Also, adding new features to these systems should be an effortless task.**

In this paper, the authors aim to design a framework for intelligent personal assistant robots. This framework can be implemented on robots that have an operating system (OS), and access to the internet. Any part of this framework is replaceable and easy to change. The robot with this framework can identify its owners and recognize their emotions and can react appropriately in real-time. For instance, after receiving a proper voice command, the robot can play a suitable song based on the owners' emotions or show body movement to convey the robot's feelings.

***Index Terms*—Human-robot interaction, Social robots, Deep learning, Computer vision, Natural language processing**

I. Introduction

In the last several years, as the result of the Corona pandemic, as well as the advancement of technology and availability of cheaper types of equipment, much attention has been paid to automating tasks using robots. As a consequence, research in the fields of robotics and machine learning has significantly accelerated. Human-robot interaction can be considered one of the sub-branches of the mentioned fields, which has received great attention. As a result, a new wave of social robots has emerged over the last few years. Social robots can interact with humans in addition to performing their tasks. These robots usually have at least one of the features of natural language processing, recognition of people's faces or actions. In recent years, in addition to trying to improve the efficiency of these systems, many efforts have been made to enhance the human-robot interaction experience by granting emotional intelligence to robots. Emotions are one of the humans' most essential characteristics, which affect many decisions. For instance, communication between humans without understanding each other's emotions will be complex and full of misunderstandings. The aforementioned propositions also apply to the relationship between humans and robots. Research has shown that the relationship between humans and robots will be more productive if the robot understands and reacts to human emotions [1]–[4]. Scientists found out humans prefer to treat robots like other human beings, and the quality of the relationship and conversation is very important. As a result, appearance is not as important as the personality and responses of the robot [5]. In addition to the mentioned cases, it has been discovered that the quality of any meeting or conversation can be determined by the distribution of emotions during the session [6]. Hence, it is possible to understand the quality of any Interaction based on emotional data.

Humans usually convey their feelings using their face, body, or voice. However, other ways, such as using electroencephalogram (EEG) or body temperature have also been used to detect human feelings. Nevertheless, facial expression recognition remains one of the most popular methods for understanding human emotions. Robots can also express emotions with their body or voice to look more natural in their interaction with humans. Paul Ekman [7]–[9], in his famous research, classified emotions into six categories of happiness, sadness, anger, surprise, fear, Neutral, and disgust. Later on, using these findings, he expanded the facial function coding system. Currently, there are generally two well-known models for classifying facial emotions, which are the discrete or categorical model and the dimensional or continuous model.

- The discrete emotional model [10]: Emotions are categorized into one of the seven basic emotions.
- The dimensional model [11]: The multidimensional model indicates the intensity of the feeling in addition to the type of feeling.

Various emotion recognition systems have been developed based on these models. Especially with the emergence of deep learning and the great success of this technique in multiple fields. Much research has been conducted in the area of facial emotion recognition in combination with deep learning, which has greatly improved the results of previous methods [12]–[15]. With the great success of these deep learning models, some of them have been implemented on robots or have been tested in unconstrained environments [16]–[20] but most of

978-1-6654-5452-0/22/$31.00 ©2022 IEEE

Fig. 1. Social robots

them were only used in the lab's controlled environments and have not been implemented in the wild.
Moreover, many efforts have been made in the field of social robots. Generally, these robots are used in the categories such as health care, service robot, education, entertainment, and transportation. For example, during the Corona pandemic, when social distancing was a common practice, and hospitals needed manpower, the presence of robots capable of performing a series of daily tasks was very helpful. The Liu robot [21], which still operates in several facilities, is a social robot developed for the said purpose. He is working and helping the patients and employees in their daily tasks in addition to interacting with them. Similarly, robots such as Nao [22] and Pepper [23] have been used in many projects with the combination of emotion recognition [24]–[30], in [31], Pepper was used for communicating with Dementia patients. Robots such as Vector [32], and Cozmo [33] as seen in figure 1 have been used as home robots for entertainment and daily use.
The problem with some of these robots is that they are primarily expensive or not designed for long-time interaction, and in most cases, there is no open-source framework [20] for future researchers to improve upon them.
In this paper, an open-source architecture for intelligent personal assistant systems is proposed, designed, and implemented on a low-cost robot. This robot uses deep models to recognize its owners' face and their emotions. Users can interact with the robot using voice commands, and the robot can give different responses based on the different feelings of its owners. Furthermore, the robot will convey its feeling with its body, and dialogue. In general, in this work, the following actions are performed:

- The major contribution of this work is a design and implementation of a simple and robust open-source framework for social robots. This framework is easy to change and implement on robots with OS and the Internet.
- The architecture is implemented on a low-cost robot. This robot can recognize the owners and their feelings and react appropriately. Owners can command the robot using their voices. Furthermore, with a program, it is possible to acquire the time and duration of emotions throughout the day for data mining and to improve future interactions.

II. Robot

This robot [34] is a low-cost robot designed using a Raspberry Pi board and ARM processor, whose complete specifications can be seen in table I.
The camera in this robot has two degrees of freedom. A servo motor can adjust the camera's angle, and a stepper motor can adjust the camera's height. This makes it possible to show interesting reactions to various events. For example, when a person is recognized as sad, by lowering the camera angle, the audience is given the feeling that the robot's head is down and it is sad.

TABLE I
Robot specifications

Specification	Values
Dimensions	28 cm $*28cm * 16\ cm$
Max velocity	20 cm/s
Weight	1kg
Motors	20 cm/s
Max velocity	70 RPM motors
Battery lifetime	3 hr
Single board computer	Raspberry pi zero w
Core processor	ARM STM32 F407
Camera	Raspberry pi camera
Camera freedom	Angular and Linear
Robot Operating System support	Yes

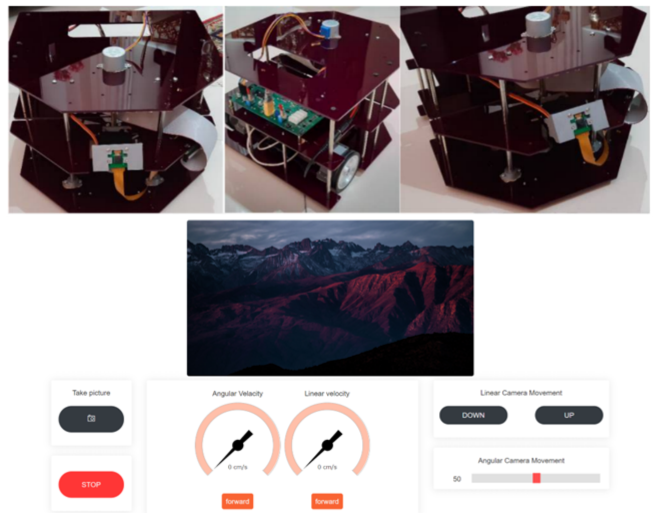

Fig. 2. The robot and the control page

III. Systems Architecture

The robot's images are sent to a server written in Python and Flask framework through a WebSocket connection. This image gets broadcasted to the clients connected to this server using the WebSocket connection. The commands for the robot's movements are received by the robot through Message Queuing Telemetry Transport(MQTT) communication, which is a widely used protocol in the Internet of Things industry. In other words, the robot listens to the broker server on a specific server topic. These topics are the commands supported by the robot, such as changing the camera angle, changing the height of the camera, and changing the speed of the motors. At first, the broker server used was the web-based test version of Mosquito, but in this project, Mosquito was installed on a local server to have a higher speed for transferring orders. The system architecture can be seen in figure 3.

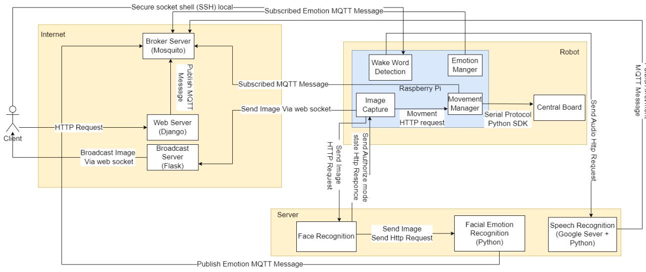

Fig. 3. System architecture

As seen in figure 2, a web page has also been designed for this robot, which can be used to communicate with the web server through an application programming interface(API) using the buttons embedded in this page. To increase the decision-making speed, some processing is done on the server. In general, processes that take time to send to the server are done on the robot, and if it is needed, the processes related to the classification of emotions and voice commands are done on the server. The architecture is shown in figure 3. The implementation in this way allows supporting commands in any language simply by replacing the voice processing server with the Google server. Furthermore, similar projects that specifically focus on audio processing can fit into this architecture.

IV. Robots Reactions

In different situations, the robot will show different predetermined reactions based on its state.

A. Waking up

The robot will wake up by saying the phrase “Hey robot(name)”. After hearing its name, the robot’s camera, which looks like a head, will be raised and starts looking for its owners around the room by spinning around. If it finds its owner, the next reaction begins; otherwise, it goes back to sleep.

B. Identification of the owner

If the robot recognizes its owner, it will stop spinning around and stares at the person’s face. Meanwhile, it commences processing the person’s emotions. After finding it out, the robot will respond to the owner with its movement and voice; for example, if any positive feeling such as happiness is sensed, the robot starts dancing with its body, and in case of negative feelings such as sadness, the robot brings its head (camera) down and asks what is wrong in a calming voice to alleviate the negative feeling. At the person’s request, the robot can play a song that is chosen based on the owner’s emotions.

If no voice commands were received or the robot could not find the owner’s face, the robot would go back to sleep and wait for the next wake-up call.

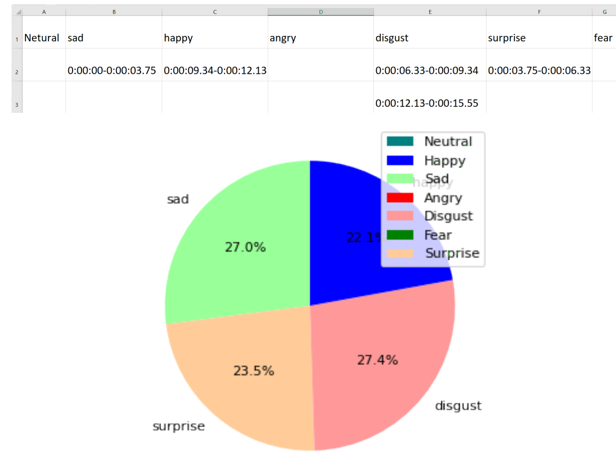

Fig. 4. Chart and timetable

C. Voice command receiving mode

After the wake-up call, the robot waits to receive a voice message for two seconds. Users can ask the robot to play a song or move its body using voice commands. In this case, by saying the word go, it moves forward. It stops when he hears the word stop. With the words left and right, it starts spinning in that direction for one second, and finally, when it hears the words up and down, it moves its camera up or down.

D. Data mining

To make future interactions more productive, the robot can record and convert the owner’s emotions and their duration into the form of a chart and an Excel file using a program that is also available to use on personal computers. The Excel file is a timetable that contains the recorded time duration for each emotion, and the chart contains the emotion distribution through the interaction. It should be mentioned that these data are only available to the owner.

V. Artificial Intelligence Models

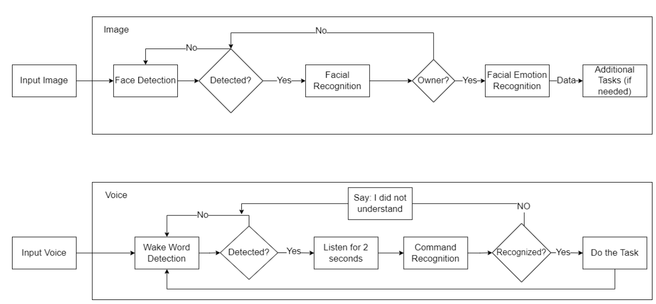

Fig. 5. Models architecture

To perform the tasks of an intelligent personal assistant, several artificial intelligence models will be used simultaneously. These models are arranged with the architecture in figure 5, each of which can be deleted or replaced with another model if needed.

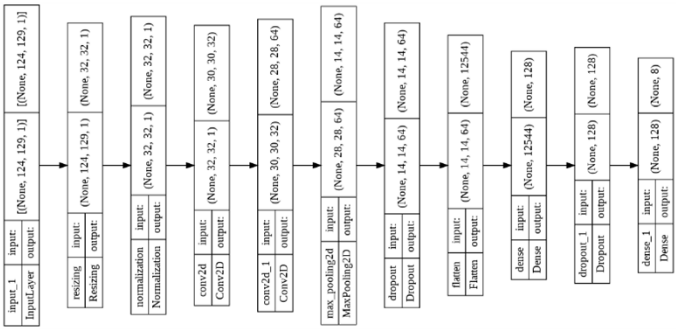

Fig. 6. Command recognition model

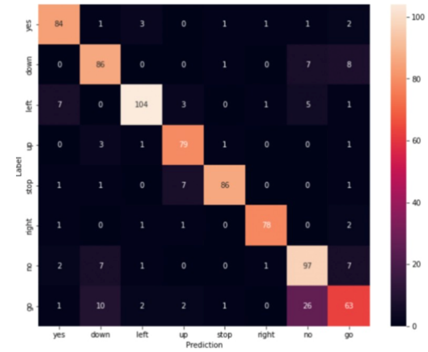

Fig. 7. Confusion matrix

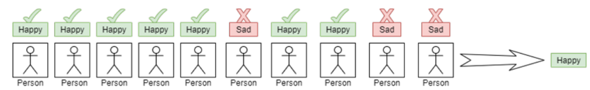

Fig. 8. Decision windows

A. Face detection model

The first step to extracting emotions from people's faces is to find their faces in pictures. In this project, the Haarcascde face detection model [35] has been used, which is a fast and high-accuracy model that can run in real-time. Furthermore, it should be mentioned that some emotion recognition models have built-in face detection models, such as Dlib [36]; in these cases, the Haarcascde model is not needed, but in the architecture, it is available to use.

B. Face recognition model

An open-source face recognition model has been used in [37]. This model is implemented using deep learning. The model has an accuracy of 99.38% on the labeled faces in the wild benchmark. To use this model, a picture of the owner should be placed in the authentication folder of the Raspberry pi with the name me.png.

C. Emotion recognition model

The deep face model [38]–[40] has been implemented as the emotion recognition model, which is lightweight face recognition and facial attribute analysis model. This model is based on the discrete emotional model, developed in Python language, and is both open source and high-accuracy, making it a great model for real-time interaction.

D. Wake word model

Picovoice API for the wake word model has been used [41], which is an on-device wake word detection powered by deep learning and can be implemented on various devices such as Raspberry pi. The phrase "Hey robot(name)" as the wake-up word was used.

E. Voice command recognition model

The model proposed in [42] has been implemented. This is a convolutional model that recognizes the input sound of the word by transforming them into a spectrogram.
For training this model, A dataset consisting of the words no, yes, go, left, right, stop, up, and down, was used. Each word has one thousand samples and eight hundred samples were used for training, hundred samples for evaluation, and the remaining hundred samples for testing. The confusion matrix obtained from the training is available in figure 7. The accuracy of the model was 85% on the test data.
It should be mentioned that the point of using these models is that they are lightweight, and any cheap processor can run them in real-time. Hence, all these models can be run simultaneously on the Raspberry Pi processor without needing any other server. (There are other models trained for mentioned tasks available on the project's Github[1])

F. Improving the emotion recognition model

Since the accuracy of each model can hamper the overall performance. Emotion detection is performed during a time window to slightly improve the robot's performance. About ten frames are processed per second, the possibility of noise in one frame can also lead to wrong decisions. Therefore, it is necessary to decide on a time window. The final result in each window is the most common emotion. In figure 8, even though the model made wrong predictions in 3 frames since the common emotion in that window was happy, the model chose the happy emotion for the overall result of that window.

VI. Experimental Result

Table II shows the results (acquired by testing around 20 samples for every class) of implementing these models on the robot. All the models are very robust, even in different scenarios. However, the Deep face model does not perform greatly in some classes, probably because of the lack of data during training or even; hence, the model is under fitted in that class. Furthermore, in general, this model's accuracy depends on the lighting of the room, which to handle this issue, gamma correction methods have been tried. However, the system generally works solid, and in most scenarios, it is fast and reliable.

[1]Link to the repository: https://github.com/MohammadJRanjbar/Social-robots-an-open-source-framework-for-personal-assistant-robots

TABLE II
EXPERIMENTAL RESULT

Models	Accuracy
Face detection	95%
Face recognition	95%
Emotion recognition	Happy: 95% Neutral: 95% Sad: 50% Disgust: 0% Surprised: 40% Angry: 70% Fear: 60%
Wake word detection	90%
Voice command	Stop: 95% Go: 80% Right: 90% Left: 85% Up: 90% Down: 80% Yes: 85% No: 80%

VII. CONCLUSION AND FUTURE WORK

In conclusion, an open-source framework for social robots has been proposed. This framework has been implemented on a low-cost robot, which performs well in various experiments as a personal assistant. This system uses several artificial intelligence models to recognize the owners, their feelings, and voice commands. These models work simultaneously in real-time and the robot can react appropriately based on the owner's emotions. Users can ask the robot to play a suitable song or move around.

The reason for the excellence of this framework is its simplicity of implementation on robots with basic features. Which, leads to improvement in their interaction with humans. Also, the components of this system are cost-effective and mostly open-source or widely used protocols.

In general, social robots with the abilities such as face, emotion, and voice recognition can be used in various fields, especially as personal assistants. Elderly, or physically disabled people can use these kinds of robots to ease their daily tasks. The robot is controllable by voice, and therefore, they can easily use voice commands to perform some of their chores quickly. However, the human-robot interaction field still has much to improve and expand, artificial intelligence systems still need to be improved and become more accurate. Robots active in this field need to get more features, both software-wise and hardware-wise. To propose future improvements based on the capabilities of this project, adding better models for more accurate recognition of emotions and implementing real-time speech emotion recognition would hugely impact the quality. Furthermore, adding more commands to make the robot more useful and connecting it to platforms such as Spotify or Soundcloud to make it easier to access different music would be helpful.

ACKNOWLEDGMENT

The authors would like to thank Mr. Amir Hossein Sheikh for providing the robot and helping to implement parts of this project.

REFERENCES

[1] McColl, D., Hong, A., Hatakeyama, N. et al. "A Survey of Autonomous Human Affect Detection Methods for Social Robots Engaged in Natural HRI". J Intell Robot Syst 82, 2016, 101–133, https://doi.org/10.1007/s10846-015-0259-2

[2] Cavallo, F., Semeraro, F., Fiorini, L., Magyar, G., Sinčák, P., and Dario, P. "Emotion modeling for social robotics applications: a review". J. Bionic Eng. 15, 2018, 185–203. DOI: 10.1007/s42235-018-0015-y

[3] Eva. "Guidelines for designing computational models of emotions". International Journal of Synthetic Emotions (IJSE), 2011. 2(1):26–79.

[4] H. R. M. Pelikan, M. Broth and L. Keevallik, ""Are You Sad, Cozmo?" How Humans Make Sense of a Home Robot's Emotion Displays," 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2020, pp. 461-470.

[5] S. Kiesler, J. Goetz, "Mental models and cooperation with robotic assistants", in Proceedings of CHI, 2002.

[6] Rößler, Jannik, Sun, Jiachen, and Gloor, Peter. "Reducing videoconferencing fatigue through facial emotion recognition". Future Internet, 13(5), 2021.

[7] Ekman, P., Friesen, W. V. "Facial Action Coding System (FACS)" [Database record]. APA PsycTests. 1978, https://doi.org/10.1037/t27734-000

[8] Ekman, P. "Are there basic emotions? Psychological Review", 99(3): 550–553. 1992.

[9] Ekman, P. "Basic emotions". In T. Dalgleish M.J. Power (Eds.), Handbook of cognition and emotion, 1999, (p. 45–60). New York: John Wiley Sons.

[10] Ekman P, Friesen WV, O'Sullivan M, Chan A, Diacoyanni-Tarlatzis I, Heider K, Krause R, LeCompte WA, Pitcairn T, Ricci-Bitti PE, et al. "Universals and cultural differences in the judgments of facial expressions of emotion". J Pers Soc Psychol. 1987 Oct;53(4):712-7. doi: 10.1037//0022-3514.53.4.712. PMID: 3681648.

[11] Lang PJ. "The emotion probe. Studies of motivation and attention". Am Psychol. 1995 May;50(5):372-85. doi: 10.1037//0003-066x.50.5.372. PMID: 7762889.

[12] Lucey, Patrick, Cohn, Jeffrey F., Kanade, Takeo, Saragih, Jason, Ambadar, Zara, and Matthews, Iain. "The extended cohn¬kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression". IEEE Computer Society Conference on Computer Vision and Pattern Recognition ¬ Workshops, 2010, pp. 94–101,

[13] Khorrami, Pooya, Paine, Tom Le, and Huang, Thomas S. "Do deep neural networks learn facial action units when doing expression recognition?" 2017.

[14] J. M. Susskind, A. K. Anderson and G. E. Hinton, i. "The Toronto face database". Department of Computer Science, University of Toronto, Toronto, ON, Canada, Tech. Rep, 2010.

[15] Minaee S, Minaei M, Abdolrashidi A. "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network". Sensors. 2021; 21(9):3046. https://doi.org/10.3390/s21093046

[16] Cid F, Moreno J, Bustos P, Núñez P, "Muecas: a multi-sensor robotic head for affective human robot interaction and imitation". Sensors 2014 (Basel, Switzerland) 14:7711–7737

[17] A. Meghdari, S. B. Shouraki, A. Siamy and A. Shariati, "The real-time facial imitation by a social humanoid robot," 2016 4th International Conference on Robotics and Mechatronics (ICROM), 2016, pp. 524-529, doi: 10.1109/ICRoM.2016.7886797.

[18] N. S. Simul, N. M. Ara and M. S. Islam, "A support vector machine approach for real time vision based human robot interaction," 2016 19th International Conference on Computer and Information Technology (ICCIT), 2016, pp. 496-500, doi: 10.1109/ICCITECHN.2016.7860248.

[19] Bera A, Randhavane T, Prinja R, Kapsaskis K, Wang A, Gray K, Manocha D "The emotionally intelligent robot: improving social navigation in crowded environments". 2019 ArXiv arXiv:1903.03217

[20] Hamza Mahdi, Sami Alperen Akgun, Shahed Saleh, Kerstin Dautenhahn, "A survey on the design and evolution of social robots — Past, present, and future, Robotics and Autonomous Systems—, Volume 156, 2022, 104193, ISSN 0921-8890, https://doi.org/10.1016/j.robot.2022.104193.

[21] Miseikis J, Caroni P, Duchamp P, Gasser A, Marko R, Miseikiene N, Zwilling F, de Castelbajac C, Eicher L, Fruh M, Fruh H. "Lio-A Personal Robot Assistant for Human-Robot Interaction and Care Applications". IEEE Robot Autom Lett. 2020 Jul 7;5(4):5339-5346. DOI: 10.1109/LRA.2020.3007462. PMID: 34192136; PMCID: PMC8043496.

[22] NAO Robot. Available online: www.softbankrobotics.com/emea/en/nao (accessed on 21 August 2022).
[23] PEPPER Robot. Available online: www.softbankrobotics.com/emea/en/pepper (accessed on 21 August 2022).
[24] A Lopez-Rincon, ”A. Emotion recognition using facial expressions in children using the NAO Robot”. In Proceedings of the International Conference on Electronics, Communications and Computers (CONIELECOMP), Cholula, Mexico, 27 February–1 March 2019; pp. 146–153.
[25] Filippini, C.; Perpetuini, D.; Cardone, D.; Merla, A. ”Improving Human-Robot Interaction by Enhancing NAO Robot Awareness of Human Facial Expression”. Sensors 2021, 21, 6438.
[26] Onder, T.; Fatma, G.; Duygun, E.B.; Hatice, K. ”An Emotion Analysis Algorithm and Implementation to NAO Humanoid Robot”. Eurasia Proc. Sci. Technol. Eng. Math. EPSTEM 2017, 1, 316–330.
[27] Ramis, S.; Buades, J.M.; Perales, F.J. ”Using a Social Robot to Evaluate Facial Expressions in the Wild”. Sensors 2020, 20, 6716.
[28] Hu, M. ”Facial Emotional Recognition with Deep Learning on Pepper Robot”. Bachelor’s Thesis, Vaasan Ammattikorkeakoulu University of Applied Sciences, Vasa, Finland, 2019.
[29] Valagkouti, Iro Athina, Christos Troussas, Akrivi Krouska, Michalis Feidakis, and Cleo Sgouropoulou. 2022. ”Emotion Recognition in Human–Robot Interaction Using the NAO Robot” Computers 11, no. 5: 72. https://doi.org/10.3390/computers11050072
[30] Catherine Caudwell, Cherie Lacey, and Eduardo B. Sandoval. ”The (Ir)relevance of Robot Cuteness: An Exploratory Study of Emotionally Durable Robot Design”. In Proceedings of the 31st Australian Conference on Human-Computer-Interaction (OZCHI’19). Association for Computing Machinery, New York, NY, USA, 2019. 64–72. https://doi.org/10.1145/3369457.3369463
[31] Muneeb I. Ahmad, Mark Khordi-moody, and Katrin S. Lohan. ”Social Robot for STEM Education”. In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI ’20). Association for Computing Machinery, New York, NY, USA, 2020. 90–92. https://doi.org/10.1145/3371382.3378291
[32] A. Triantafyllidis, A. Alexiadis, D. Elmas, K. Votis, D. Tzovaras, ”A social robot-based platform for prevention of childhood obesity”, Proceedings - 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering, BIBE 2019 (2019) 914–917doi:10.1109/BIBE.2019.00171.
[33] Saminda Sundeepa Balasuriya, Laurianne Sitbon, Margot Brereton, and Stewart Koplick. 2019. ”How can social robots spark collaboration and engagement among people with intellectual disabilities? In Proceedings of the 31st Australian Conference on Human-Computer-Interaction” (OZCHI’19). Association for Computing Machinery, New York, NY, USA, 209–220. https://doi.org/10.1145/3369457.3370915
[34] A. Sheikholeslami, ”Design and implementation of a wheeled robot to adapt to the camera and provide panoramic image from several images”, 2022.
[35] P. Viola and M. Jones, ”Rapid object detection using a boosted cascade of simple features,” Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001, pp. I-I, DOI: 10.1109/CVPR.2001.990517.
[36] “A toolkit for making real-world machine learning and data analysis applications in” dlib. Accessed August 22, 2022. http://dlib.net/
[37] The world’s simplest facial recognition API for Python: https://github.com/ageitgey/face_recognition (accessed on 21 August 2022).
[38] S. I. Serengil and A. Ozpinar, ”HyperExtended LightFace: A Facial Attribute Analysis Framework,” 2021 International Conference on Engineering and Emerging Technologies (ICEET), 2021, pp. 1-4, doi: 10.1109/ICEET53442.2021.9659697.
[39] S. I. Serengil and A. Ozpinar, ”LightFace: A Hybrid Deep Face Recognition Framework,” 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 2020, pp. 1-5, doi: 10.1109/ASYU50717.2020.9259802.
[40] ”A Lightweight Face Recognition and Facial Attribute Analysis (Age, Gender, Emotion, and Race)” https://github.com/serengil/deepface (accessed on 21 August 2022).
[41] ”Picovoice Wake word platform” https://picovoice.ai/ (accessed on 21 At 2022).
[42] “Simple Audio Recognition: Recognizing Keywords: Tensorflow Core.” TensorFlow. Accessed August 22, 2022. https://www.tensorflow.org/tutorials/audio/simple_audio.