# A Visual Information-based Bidirectional Emotion Interaction Interface for Friendly and Empathic Collaborative Robots

Jordan Murphy
*School of Computing*
*Montclair State University*
Montclair, NJ, USA
murphyj37@montclair.edu

Jesse Parron
*School of Computing*
*Montclair State University*
Montclair, NJ, USA
parronj1@montclair.edu

Maxim Lyons
*School of Computing*
*Montclair State University*
Montclair, NJ, USA
lyonsm1@montclair.edu

Weitian Wang
*School of Computing*
*Montclair State University*
Montclair, NJ, USA
wangw@montclair.edu

Rui Li*
*School of Computing*
*Montclair State University*
Montclair, NJ, USA
liru@montclair.edu

2024 International Conference on Networking, Sensing and Control (ICNSC) | 979-8-3503-6522-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICNSC62968.2024.10760097

***Abstract*— Human-robot collaboration in manufacturing is an increasingly important topic in the field of robotics. However, the current collaborative robot's unresolved mechanical and stiff behaviors make its interaction with humans extremely dull, especially for an extended period. Such interaction patterns could further discourage humans from collaborating with these robots. To solve this question, this paper aims to enable collaborative robots' emotional ability in recognizing and responding to human emotions through developing a visual information-based bidirectional emotion interaction (VI-BEI) interface and testing it in manufacturing co-assembly tasks. The developed interaction interface enables the robot to not only recognize human emotions visually but also provide artificial emotion feedback via 3D simulation technology which has flexible and quick prototypes, customization, and upgrading advantages compared to hardware design. Specifically, this paper introduces (1) the development of a 3D digital human interface that not only monitors human facial expressions but also produces artificial emotion feedback. (2) integrating the 3D digital human interface to enable a collaborative robot to express real-time emotions in addition to performing actions during co-assembly tasks, facilitating a friendly collaboration process. (3) validation experiments and analysis to evaluate the effectiveness and performance of the updated collaborative robot with facial expressions through real-world assembly tasks. The experimental results and analysis demonstrate the effectiveness and advantages of the current system, as well as guide the future improvement of the developed collaborative robots to be more empathic and friendly.**

***Keywords*— Robotics, Bidirectional Interaction, Human-Robot Interaction, Human/Computer Interface***

## I. INTRODUCTION

Human-robot collaboration [1]–[4] is an increasingly important topic in the field of robotics. Collaborative robots are typically designed to work closely with humans in a shared space and have been widely used in manufacturing industries such as automotive, food, and pharmaceutical production. The purpose of using collaborative robots to assist humans in dull, repetitive, and dangerous working tasks is to improve manufacturing efficiency and productivity [1]–[3], [5], [6]. Though current collaborative robots can assist in improving productivity, the unresolved mechanical and stiff behaviors of such collaborative robots make their communication and interaction with humans extremely dull, especially for an extended period [7]. Such mechanical and stiff interaction patterns discourage human's willingness to work with collaborative robots. It further negatively impacts user acceptance and the wide application of collaborative robots in manufacturing areas.

To solve these questions and be inspired by human-human collaboration, this paper aims to enable robots' basic emotional abilities in recognizing and responding to human emotions by a developed visual information-based bidirectional emotion interaction interface for collaborative robots, and uses manufacturing co-assembly tasks as working scenarios to test the developed system. The developed emotion based bidirectional interaction interface enables the robot not only to recognize human emotions visually but also to provide artificial emotion feedback via 3D simulation technology. The benefit of using 3D simulation technologies lies in their flexible and quick prototypes, customization, and upgrading advantages compared to hardware. This configuration further benefits the study of effective facial expressions in a friendly human-robot interaction. Specifically, our work includes the following three parts. First, the development of a 3D digital human that not only monitors human facial expressions but also produces artificial emotion feedback. Second, integrating the 3D digital human interface enables a collaborative manufacturing robot to express real-time emotions in addition to performing actions during co-assembly tasks. It facilitates a friendly collaboration process. Third, validation experiments and analysis to evaluate the effectiveness and performance of the updated collaborative robot with facial expressions through real-world assembly tasks. The experimental results and analysis demonstrate the

*Corresponding author: Dr. Rui Li (liru@montclair.edu).

effectiveness and advantages of the current system, as well as guide the future improvement of the developed collaborative robots to be more empathic and friendly.

The contributions of our work can be summarized as follows: (1) Development of a visual system that allows the robot to monitor and promptly recognize human facial expressions. This enables the robot to perceive different human emotions visually, similar to humans. (2) Development and integration of a 3D digital human to enable the collaborative robot to generate artificial emotions in addition to collaborative actions. This further enhances an empathic interaction process. (3) Integration and testing of the developed robot with emotional expression ability into real-world co-assembly tasks. This enables exploration of the actual applications of the developed robotics system and provides a guide on the future development of empathic and friendly robots. (4) Combining the flexible and rapid prototypes, customization, and upgrading advantages of 3D technology with robotics hardware to enhance an easier study of emotion factors that can improve human-robot collaboration.

## II. Related Works

### A. Visual Information Based Human Emotion Identification

In the study of human emotion identification, the initial step involves representing different emotions in forms that are analyzable and computable [8]. One of the most important approaches is the discrete representation model, which classifies emotions into distinct categories. According to Ekman's basic emotion theory [9], human emotions can be classified as six fundamental emotions: fear, anger, joy, sadness, disgust, and surprise. Similarly, Tomkins et al. [10] categorized emotions into seven types based on varying intensities. This includes interest-excitement, surprise-startle, enjoyment-joy, anger-rage, distress-anguish, shame-humiliation, and fear-terror.

After representing human emotions in a form that can be analyzed and computed, human emotion identification methods can further calculate different emotions. For human emotion identification, camera-based emotion recognition is one of the popular methods due to it is flexible and easy to implement. For camera-based emotion recognition, two main features have been used for emotion recognition: facial expressions [11]–[14] and body gestures [15], [16]. For facial expressions, Teixeira et al. [17] constructed a spatiotemporal convolutional neural network for predicting continuous emotional values of valence and arousal using facial expression data. Faria et al. [18] extracted simple features of facial landmarks distances and angles for discrete emotion identification based on facial expression recognition. Then, a dynamic probabilistic classification model was built to output seven discrete emotional statuses (angry, fearful, disgusted, happy, sad, surprised, and neutral). Other research uses human body gestures for emotion identification. For example, Sun et al. [19] developed a long short-term memory recurrent neural network (LSTM_RNN) model to calculate human emotions based on body pose. Yan et al. [20] integrated features from facial expressions and body gestures for motion recognition. Piana et al. [21] use three-dimensional motion data captured by Kinect for human emotion identification. These camera-based methods are easy to use, and further enhance fast and real-time emotion recognition for collaborative robots.

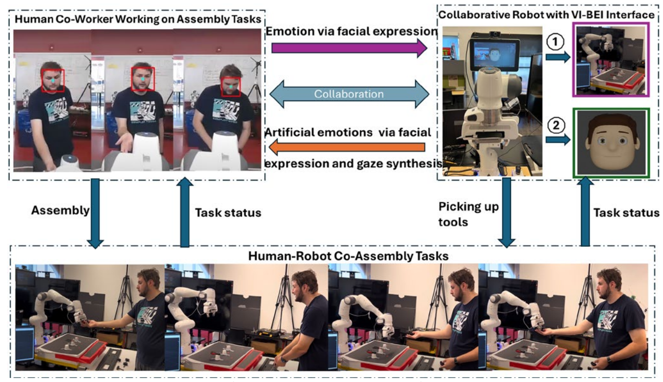

Figure 1. System overview

### B. Facial Expression Based Artificial Emotion Synthesis and Eye Tracking

Facial expression and eye tracking are very important in human-human communication. Humans are good at expressing their emotions via facial expressions and showing their interest by gazing into other's eyes. Inspired by this, multiple research projects have been done for synthesizing facial expressions and enabling eye-tracking functions for robots to enhance a friendly human-robot interaction process. Currently, there are two main methods to realize this: hardware/mechanical simulation and software simulation.

For the hardware/mechanical simulation, the robot's face is designed with groups of mechanical actuators and soft materials that simulate the muscle movements and skin appearance. The robotics facial expressions are further simulated by controlling these actuators. Classical examples of such robots include Sophia robot [22] and Affetto robot [23]. However, the uncanny valley effect [24]–[27] is always the main challenge that makes such robots very hard to be widely accepted by users.

An alternative approach is to use software simulation technology to mimic facial expressions. In this method, virtual human models can be built and animated for integration with the robot system. For example, the Baxter robot [28] is a very classic example that uses 2D animation to mimic a simple face with eyebrows and eyes to enhance the interaction ability of a manufacturing robot. Nao robot [29]–[31] is another example that uses light color changes to simulate some emotions via eyes. Compared to hardware simulation, the software simulation-based method has the advantages of flexible and quick prototypes, customization, and upgrading. The uncanny valley effect can be easily and quickly noticed and corrected in this approach by adjusting the simulation results. Hence, this paper chooses to use 3D simulation technology to synthesize artificial emotions for the robot.

## III. System Overview

To enable collaborative robots' basic emotional ability, this paper aims to integrate visual information-based bidirectional emotion interaction (VI-BEI) in the process of human-robot collaboration. Manufacturing co-assembly is used as a working context to assess the developed system. Fig. 1 shows the system overview demonstrating how the developed VI-BEI interface contributes to the human-robot co-assembly tasks.

The human-robot co-assembly includes three important components: the human worker, the collaborative robot, and assembly tasks. The collaborative robot designed in this paper includes two parts: a Franka Emika Panda robot, and a 3D simulated virtual head with the ability to express synthesized emotions. In the collaborative assembly tasks, the robot will be responsible for picking up assembly tools in the necessary order and delivering them to the human. The human will be responsible for finishing the rest of the assembly tasks. In addition to delivering the tool to the human, the robot's interaction abilities are further enhanced by the developed VI-BEI interface. This interface also enables the robot to monitor human emotions through visual information streaming via a webcam. Based on the integrated facial expression recognition algorithm, the robot will be able to identify human emotions. Moreover, the robot also can express its own artificial emotions in the form of 3D synthesized visual feedback via the VI-BEI interface. Specifically, this feedback includes 3D facial expressions based on recognized human emotions as well as eye gaze synthesis for an empathic and friendly interaction with humans.

## IV. METHODS AND MATHEMATICAL MODELS

### A. Facial Expression Based Emotion Recognition

To realize the goal in this paper, the robot needs to be able to see and recognize human emotions. This paper integrated the Viola-Jones object detection framework [32] for quick human face detection through a web camera. The integral images and Haar-like feature-based AdaBoost learning enable the Viola-Jones object detection framework to perform quick detection tasks that satisfy our experimental needs.

Once a face is detected, the system needs to further identify what emotion the human could be feeling at the time. For facial expression identification, this system implements a lightweight DeepFace framework [33], [34] integrated with the Facenet512 recognition model. The Facenet512 has the best recognition accuracy compared to the other recognition models provided in the DeepFace framework. The Facenet512 is built using a convolutional neural networks (CNN) architecture. CNN is a neural network with multiple layers that is trained to predict accurate results in machine learning problems. The input images first go through the convolution layer, where they pass through multiple convolution filters to find certain features in the image. The equation for the convolution of one pixel in the next layer is as shown:

$$net(t,j) = (x * w)[t,j] = \sum_m \sum_n = x[m,n]w[t-m, j-n], \tag{1}$$

where $net(t,j)$ is the output on the following layer, $x$ is the input image and $w$ is the filter matrix. From there, the image goes through a nonlinearity layer to adjust the output from the previous layer. Using the Rectified Linear Unit (ReLU), to further improve the training of the neural network by conventionally being used as an activation function, the output is either saturated or limited depending on the situation, using the following equations:

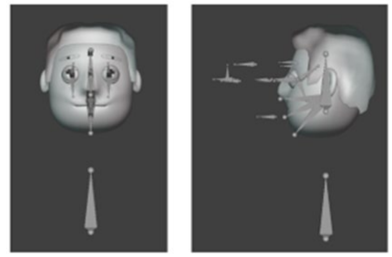

Figure 2. Example of the developed armature for controlling facial animation.

$$ReLU(x) = \max(0, x), \tag{2}$$

$$\frac{d}{dx}(x) = \{1 \; if \; x > 0; 0 \; otherwise\}, \tag{3}$$

To create a fully connected layer, the output from the nonlinearity layer must go through the pooling layer, basically similar to reducing the resolution.

### B. 3D Digital Human-Based Artificial Emotion Synthesis

To give the user a method of communication and interaction with the system, a 3D interactive digital human is created. It is controlled based on Blender's armature system where the assigned bones can be moved around in the 3D space to affect the mesh of the head model. The mathematical representation of the bones in the armature can be described as a set. The set consists of multiple sets for specific controls and can be represented as

$$F = \{b_i^{head}, b_j^{mouth}, b_k^{eyes}, b_l^{eyelids}, b_m^{eyebrows}\}, \tag{4}$$

where $b_i^{head}$, $i \in M$ denotes the $i$th controlling bone for the head region; $b_j^{mouth}$, $j \in N$ denotes the $j$th controlling bone of mouth region; $b_k^{eyes}$, $k \in \mathrm{K}$ represents the $k$th controlling bone of eye regions; $b_l^{eyelids}$, $l \in \mathrm{T}$ indicates the $l$th controlling bone of eyelid regions; and $b_m^{eyebrows}$, $m \in N$ denotes the $m$th controlling bone for eyebrow regions.

Each control region in the set $F$ consists of multiple sets of bones $b_\mu^\nu$, $\nu \in \{head, mouth, eyes, eyelids, eyebrows\}$, and $\mu$ is the index of the bone. The location of each bone is dependent on its values in the 3D space and any transformations influenced by it can be calculated by the following equation:

$$b_{\mu'}^{\nu} = \boldsymbol{H} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \tag{5}$$

where the (x, y, z) indicates the 3D coordinates of the $b_\mu^\nu$, $\boldsymbol{H} = \Theta_T \cdot \Psi_S \cdot \Pi_R$, where $\Theta_T$ is the translation matrix, $\Psi_s$ is the scaling matrix, and $\Pi_R$ is the rotation matrix.

To bring life to the interactive digital human, some poses and animation needed to be made so the virtual face could express emotions as a response to the user. To do so, the 3D armature was used to manipulate the various parts of the face. The armature consists of a set of bones that are used for controlling the animation. By changing the location and rotation of the armature in the areas of head, mouth, eyes, eyelids, and eyebrows, different facial expressions can be crafted by having images of human faces or a mirror handy to use as reference. Simply bending the cheeks up or down can help create more complex facial reactions. The location of the eyebrows and the extent to which the eyelids are open can create different confused or annoyed expressions. The overall head angle can also change how a facial expression can be perceived. Fig. 2 shows the designed armature structure inside the 3D digital human, where each visible bone has a different level of control over the mesh of the 3D model. Some bones are specifically made to control other bones as well, like a control bone for the eyes, eyelids, lips, and the head as a whole.

### C. *Eye Gaze Synthesis for the 3D Digital Human*

Eye contact plays an important role in communication. It helps people to establish trust, express interest, and convey emotions and intentions. Inspired and motivated by this, we developed functions to enable the robot's eye contact with humans for friendly and empathic human-robot interaction via 3D eye gaze synthesis for the integrated 3D digital human.

This is realized by enabling the eyes of the robot to trace a key point $K^{gaze}(t) = (x^{gaze}(t), y^{gaze}(t))$ on the human face in the 2D camera view. The key point is located on the human nose and near the human eyes. With human moves, this key point will move as well. To measure the extent of this movement, $V^{gaze}(t) = \frac{K^{gaze}(t)}{L}$ , where L is the dimension(s) of the webcam feed window. The eye gaze direction $G^{robot}(t)$ of the robot then can be calculated based on movements of the key point $V^{gaze}(t)$:

$$G^{robot}(t) = \gamma * (b_{k,max}^{eye} - b_{k,min}^{eye}) * V^{gaze}(t) - b_{k,min}^{eyes}, \quad (6)$$

where $\gamma$ is the ratio factor for correcting directions, $b_{k,max}^{eye}$ and $b_{k,min}^{eye}$ are the maximum and minimum movement extent of the controlling bones of the eye regions indicated in Equation (4). The above Equation (6) builds a map between the movements of human eyes and head defined by $K^{gaze}(t)$ and changes in the robot's eyes gaze defined by $G^{robot}(t)$.

## V. EXPERIMENTS

### A. *Experimental Setup*

The developed robotics system runs two parallel programs for the VI-BEI interface and robot arm control. The VI-BEI interface is developed by Blender and Python with libraries installed such as math, mathutils, OpenCV, and Deepface. This setup allows the robot not only to recognize human emotions visually but also to provide artificial emotion feedback via 3D simulation technology.

The collaborative robot used in the experiments was a Franka Emika Panda robot which is a seven-axis robot arm with a 6.6lb payload and 85cm of reach. The robot arm used in this experiment is mounted on a workbench to collaborate with a human worker in assembling a TV monitor stand. Three common tools were used for finishing the assembly task: a screwdriver, pliers, and an Allen wrench.

To test how well the designed robot can react to the human in the collaborative assembly task of putting together part of a TV stand, the following three scenarios were proposed: (1) the robot hands each tool promptly to complete the task at hand, (2) the robot delays in handing the tools which irritate the human and can cause the task to fail, (3) the robot intentionally hands the incorrect tools, leading to the human becoming angry and the task failing.

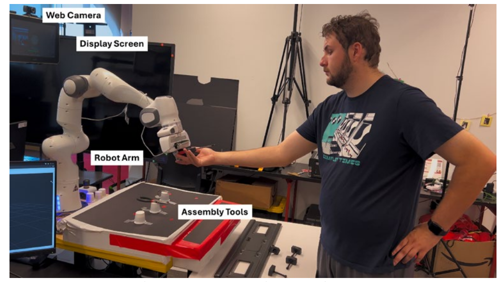

Figure 3. Experimental setup

To achieve a realistic result, we set up a timer for limiting each task based on how long they would normally take without assistance. Fig. 3 displays an example of an experimental setup which includes a web camera for human face monitoring, a screen for visual feedback display (recognized human emotions and synthesized robot emotions) display, a collaborative robot arm, and the assembly task.

### B. *Experimental Results*

The human works with the robot to complete the assembly task of putting together part of a TV stand. The robot hands the tools the human needs in a certain order. The designed collaboration starts with the human waiting for the tool they need to be passed over by the robot. The robot then picks up the requested tool and gives it to the human. The human then completes the current part of the assembly task and responds to the interactive digital human. Later, the tool is returned to its starting position. This cycle will repeat until the assembly task is fully completed. This collaborative pattern is repeated for each proposed scenario, where at first the tools are passed along promptly, then it experiences some delays and issues with handing the tools in the remaining scenarios.

Fig. 4 shows examples of recorded results. For each picture, the recognized real-time emotion is highlighted in the webcam feed on the left and the facial expression feedback of the robot depending on the average recognized emotion is shown on the right. The eye gaze synthesis is based on the position of the center of the recognized human's face (Equation (6)) in the camera feed. In each instance, the human worker starts patiently waiting for the robot to hand them the necessary tools for the job. In Fig. 4(a1) to Fig. 4(a3), the robot hands the human worker a tool they need promptly, allowing them to complete their task and be ready to repeat the process in Fig. 4(a4) to Fig. 4(a5). During this period, the robot gathered that the human worker was mostly happy with the help it offered, and it returned happy facial expressions as a response to the human through the 3D digital human. For Fig. 4(b1) to Fig. 4(b3), the robot was delayed in handling the tool and the human worker fell behind schedule because of it. This resulted in the human being upset about the delay in their task and impatient to finish, as seen in Fig. 4(b4) to Fig. 4(b5). The robot saw this and first returned a comforting smile trying to return a friendly

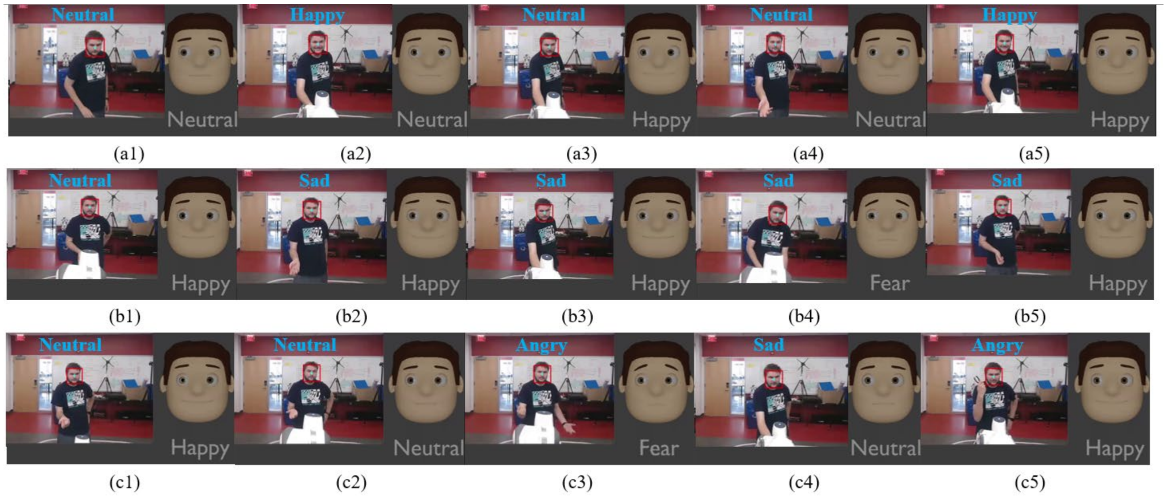


Figure 4. Three experiment scenarios where (a) the robot hands each tool in a timely manner to complete the task at hand and resulting in a happy response from the human, (b) the robot delays in handing the tools which irritates the human and can cause the task to fail, (c) the robot intentionally hands the incorrect tools, leading to the human becoming angry and the task failing.

expression. While the human was working on their task, the robot noted that the human felt fearful, and worried about not finishing on time. It returned a similar fearful face, worried if the human was okay as seen in Fig. 4(b4). This enables the robot to show empathy. In Fig 4(c1) to Fig 4(c2), the robot was heavily delayed in grabbing the tool, and ended up grabbing the wrong tool, which irritated the human as seen in Fig. 4(c3). Fig. 4(c3) also shows that the robot displayed a fearful facial expression when recognizing the human was angry. This ended up repeating, as seen in Fig 4(c4) to Fig. 4(c5), each time the robot was worried and tried to return mostly comforting facial expressions.

### C. *Experimental Analysis and Comparison*

The experimental results have demonstrated the effectiveness of our developed system in recognizing human facial emotions and making corresponding responses via the integrated visual information-based bidirectional emotion interaction (VI-BEI) interface. The performance of the developed VI-BEI system is shown in Table 1. The emotion recognition speed is 0.01s, facial emotion recognition accuracy is 99.65%. The fastest speed of the virtual feedback via 3D animation simulation is 0.2 time/s depending on the current configuration (Intel® Core™ i5-9300H @ 2.4GHz, Intel® UHD Graphics 630, NVIDIA GeForce GTX 1650, 12GB RAM) of the computer we used. We also believe the current version of this proposed system can be further improved by upgrading computational hardware. Due to the camera being placed in a fixed place, sometimes, it cannot read the human's emotions when the human is turning away. We believe this issue can be improved by introducing a multimodal human emotion recognition method.

Table 1. System Performance

| Metrics | Performance |
|---|---|
| Emotion recognition speed | 0.01s |
| Emotion recognition accuracy | 99.65%[1] |
| Virtual feedback rate | 0.2 time/s |

[1] *Performance of the Facenet512*

Compared to the manufacturing robot Baxter [28] which only can generate very simple facial expressions, our system uses a 3D digital human who can make complex facial expressions. Moreover, Baxter cannot communicate with the person it is assisting, it is limited to where it is placed and the task it is trained to do. Since few manufacturing robots consider emotion, we further compared our system with other robots designed for social service. For example, the NAO robot can communicate verbally, move around, and recognize people and objects, while it is currently limited to expressing simple emotions via changing the light color of eyes and some body gestures. In addition, NAO is only a social companion robot and cannot lift heavy parts compared to a heavy-duty arm like the Franka Emika used in the experiment [35]. There is also the Sophia robot [22], one that can be used in a more social setting and looks very similar to a human being. This similarity could make some people uncomfortable as it is pushing into the uncanny valley, where people's positive look on a robot can shift to repulsion the more the robot looks and acts like a human [36]. Our developed interactive 3D digital human has been designed to avoid uncanny valley effects by a cartoonish character.

## VI. Conclusion & Future Work

In this paper, a visual information-based bidirectional emotion interaction (VI-BEI) interface has been developed and integrated into the traditional collaborative robot system to enable the robot's basic emotional ability in recognizing and responding to human emotions through visual information. Three manufacturing co-assembly scenarios have been used as the collaborative context to test the effectiveness of the developed interactive interface in this paper. The experimental results have demonstrated the developed robot's ability to promptly recognize human workers' emotions as well as express its own artificial emotions for a friendly and empathic collaboration process. Future works will be focused on improving emotion recognition, especially for when a person looks away as well as adding voice control over the robot into the system. Another future work will be improving the current 3D facial expression synthesis system for more accurate facial expressions. All of this is to enhance a more friendly and empathic human-robot interaction.

## ACKNOWLEDGMENT

This work is supported by the National Science Foundation under Grant CMMI-2301678 and Grant CMMI-2338767.

## REFERENCES


[1] W. Wang, R. Li, Z. M. Diekel, Y. Chen, Z. Zhang, and Y. Jia, "Controlling object hand-over in human-robot collaboration via natural wearable sensing," *IEEE Trans Hum Mach Syst*, vol. 49, no. 1, pp. 59–71, 2019.

[2] W. Wang, R. Li, Y. Chen, Z. M. Diekel, and Y. Jia, "Facilitating human-robot collaborative tasks by teaching-learning-collaboration from human demonstrations," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 640–653, 2019.

[3] W. Wang, Y. Chen, R. Li, Z. Zhang, V. Krovi, and Y. Jia, "Human-robot collaboration for advanced manufacturing by learning from multi-modal human demonstrations," in *Manufacturing in the Era of 4th Industrial Revolution*, World Scientific Publishing Co., 2021, pp. 87–116.

[4] S. A. Green, M. Billinghurst, X. Chen, and J. G. Chase, "Human-robot collaboration: a literature review and augmented reality approach in design," *Int J Adv Robot Syst*, vol. 5, no. 1, pp. 1–18, 2008.

[5] W. Wang, R. Li, Z. M. Diekel, Y. Chen, Z. Zhang, and Y. Jia, "Controlling object hand-over in human-robot collaboration via natural wearable sensing," *IEEE Trans Hum Mach Syst*, vol. 49, no. 1, pp. 59–71, 2018.

[6] C. J. Conti, A. S. Varde, and W. Wang, "Human-robot collaboration with commonsense reasoning in smart manufacturing contexts," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 1784–1797, 2022.

[7] R. Li, and W. Wang, "Augmenting the communication naturalness via a 3D audio-visual virtual agent for collaborative robots," in *IEEE International Conference on Big Data, Big Data*, IEEE, 2021, pp. 5944–5946.

[8] M. Spezialetti, G. Placidi, and S. Rossi, "Emotion recognition for human-robot Interaction: recent advances and future perspectives," *Front Robot AI*, vol. 7, no. December, pp. 1–11, 2020.

[9] P. Ekman, "An argument for basic emotions," *Cogn Emot*, vol. 6, no. 3–4, pp. 169–200, 1992, doi: 10.1080/02699939208411068.

[10] S. S. Tomkins, *Affect imagery consciousness: the complete edition: two volumes*. Springer publishing company, 2008.

[11] L. Chen, W. Su, M. Wu, and W. Pedrycz, "A fuzzy deep neural network with sparse autoencoder for emotional intention understanding in human-robot interaction," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 7, pp. 1252–1264, 2020.

[12] T. Sai Teja, and M. K. Mariam Bee, "Facial expression recognition in video with multiple feature fusion," *Int J Eng Adv Technol*, vol. 8, no. 6, pp. 362–365, 2019.

[13] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "Discriminant functional learning of color features for the recognition of facial action units and their intensities," *IEEE Trans Pattern Anal Mach Intell*, vol. 41, no. 12, pp. 2835–2845, 2019.

[14] L. Chen, M. Zhou, W. Su, M. Wu, J. She, and K. Hirota, "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction," *Inf Sci (N Y)*, vol. 428, pp. 49–61, 2018.

[15] J. Yan, W. Zheng, M. Xin, and J. Yan, "Integrating facial expression and body gesture in videos for emotion recognition," *IEICE Trans Inf Syst*, vol. E97-D, no. 3, pp. 610–613, 2014, doi: 10.1587/transinf.E97.D.610.

[16] S. Piana, A. Staglianó, F. Odone, and A. Camurri, "Adaptive body gesture representation for automatic emotion recognition," *ACM Trans Interact Intell Syst*, vol. 6, no. 1, pp. 1–31, 2016, doi: 10.1145/2818740.

[17] T. Teixeira, É. Granger, and A. L. Koerich, "Continuous emotion recognition with spatiotemporal convolutional neural networks," *Applied Sciences (Switzerland)*, vol. 11, no. 24, 2021, doi: 10.3390/app112411738.

[18] D. R. Faria, M. Vieira, and F. C. C. Faria, "Towards the development of affective facial expression recognition for human-robot interaction," *ACM International Conference Proceeding Series*, vol. Part F1285, pp. 300–304, 2017, doi: 10.1145/3056540.3076199.

[19] M. Sun, Y. Mou, H. Xie, M. Xia, M. Wong, and X. Ma, "Estimating emotional intensity from body poses for human-robot interaction," 2019.

[20] J. Yan, W. Zheng, M. Xin, and J. Yan, "Integrating facial expression and body gesture in videos for emotion recognition," *IEICE Trans Inf Syst*, vol. E97-D, no. 3, pp. 610–613, 2014, doi: 10.1587/transinf.E97.D.610.

[21] S. Piana, A. Staglianó, F. Odone, and A. Camurri, "Adaptive body gesture representation for automatic emotion recognition," *ACM Trans Interact Intell Syst*, vol. 6, no. 1, pp. 1–31, 2016, doi: 10.1145/2818740.

[22] T. Riccio, "Sophia robot," *TDR: The Drama Review*, vol. 65, no. 3, pp. 42–77, 2021, doi: 10.1017/s1054204321000319.

[23] H. Ishihara, Y. Yoshikawa, and M. Asada, "Realistic Child Robot 'Affetto' for Understanding the Caregiver-Child Attachment Relationship that Guides the Child Development," in *IEEE International Conference on Development and Learning (ICDL))*, IEEE, 2011, pp. 1–5.

[24] M. Mori, K. MacDorman, and N. Kageki, "The uncanny valley," *IEEE Robot Autom Mag*, vol. 19, no. 2, pp. 98–100, 2012, doi: 10.1109/mra.2012.2192811.

[25] T. Geller, "Overcoming the uncanny valley," *IEEE Comput Graph Appl*, vol. 28, no. 4, pp. 11–17, 2008, doi: 10.1109/MCG.2008.79.

[26] M. A. Bassi, M. A. Lopez, L. Confalone, R. M. Gaudio, L. Lombardo, and D. Lauritano, "Avoiding the uncanny valley in virtual character design," *Interactions*, vol. 25, no. 5, pp. 45–49, 2018.

[27] M. Destephe, M. Brandao, T. Kishi, M. Zecca, K. Hashimoto, and A. Takanishi, "Walking in the uncanny valley: importance of the attractiveness on the acceptance of a robot as a working partner," *Front Psychol*, vol. 6, 2015, doi: 10.3389/fpsyg.2015.00204.

[28] N. T. Fitter, and K. J. Kuchenbecker, "Designing and assessing expressive open-source faces for the Baxter robot," in *International Conference on Social Robotics*, 2016, pp. 340–350.

[29] M. S. Erden, "Emotional postures for the humanoid-robot Nao," *Int J Soc Robot*, vol. 5, no. 4, pp. 441–456, 2013, doi: 10.1007/s12369-013-0200-4.

[30] R. Andreasson, B. Alenljung, E. Billing, and R. Lowe, "Affective touch in human-robot interaction: conveying emotion to the nao robot," *Int J Soc Robot*, vol. 10, no. 4, pp. 473–491, 2018, doi: 10.1007/s12369-017-0446-3.

[31] A. Tapus, A. Peca, A. Aly, C. Pop, L. Jisa, S. Pintea, A. Rusu, and D. David, "Children with Autism Social Engagement in Interaction with Nao , an Imitative Robot - A Series of Single Case Experiments," *Interact Stud*, vol. 3, no. 13, pp. 315–347, 2012.

[32] P. Viola, and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features."

[33] S. Serengil, and A. Özpınar, "A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules," *Bilişim Teknolojileri Dergisi*, vol. 17, no. 2, pp. 95–107, 2024.

[34] S. I. Serengil, and A. Ozpinar, "Lightface: A hybrid deep face recognition framework," in *2020 innovations in intelligent systems and applications conference (ASYU)*, IEEE, 2020, pp. 1–5.

[35] S. Haddadin, S. Parusel, L. Johannsmeier, S. Golz, S. Gabl, F. Walch, M. Sabaghian, C. Jahne, L. Hausperger, and S. Haddadin, "The Franka Emika Robot: A Reference Platform for Robotics Research and Education," *IEEE Robot Autom Mag*, vol. 29, no. 2, 2022, doi: 10.1109/MRA.2021.3138382.

[36] C. Thepsoonthorn, K. ichiro Ogawa, and Y. Miyake, "The Exploration of the Uncanny Valley from the Viewpoint of the Robot's Nonverbal Behaviour," *Int J Soc Robot*, vol. 13, no. 6, 2021, doi: 10.1007/s12369-020-00726-w.