Facial recognition for coexistence In the home

Diego Carrillo Arroyo *, Jose Serrano Quesada †, Andrik Solano Jimenez ‡ y Olman Acuña Quesada §
Escuela de Computación San Carlos, Instituto Tecnológico de Costa Rica
Email: *dcgreenforest@estudiantec.cr, †2020042060@estudiantec.cr, ‡asolanojim@estudiantec.cr, §olman02@estudiantec.cr

***Abstract*—In the area of smart homes, where various technologies converge to enhance the quality of daily life, understanding the emotional states of inhabitants becomes crucial for harmonious coexistence. This research delves into the development and implementation of a cutting-edge facial recognition system dedicated to emotion detection for in-home environments. By utilizing advanced camera systems installed in the household, the proposed solution is capable of identifying and analyzing the facial expressions of family members, subsequently discerning their emotional state. Once detected, these emotional states are then transmitted to other members of the household via smartphone notifications, aiming to foster an understanding and supportive living environment. This system, thereby, holds potential in not only improving interpersonal dynamics within the family but also in integrating emotion-aware smart home functionalities, such as ambient adjustment based on mood. The results indicate promising accuracy in emotion detection and timely notification delivery, suggesting myriad applications in future smart home configurations.**

***Keywords*—Facial Recognition, AI, Smart Homes.**

I. Introduction

In this digital age, homes are increasingly an integrated system of devices connected to each other, so the incorporation of emotion-sensitive technologies offers a novel approach to improve home harmony . Because as family members interact in daily life, understanding and responding to each other's emotions becomes an important point in fostering coexistence. This project seeks to harness the capabilities of facial recognition technology to create a bridge of emotional understanding within the family unit, in the context of our increasingly interconnected living spaces.

Facial recognition, a subset of computer vision that has seen significant advances in recent years, driven by the evolution of deep learning algorithms and increasing computational power [7]. In which it has been used mainly for security and identity verification purposes. Extending this technology to discern emotions adds a new dimension to its potential applications. The human face has countless expressions that provide a lot of information, so when interpreted accurately, it can give insight into a person's emotional state [6].

Additionally, recognizing a family member's mood can help prevent conflict, provide support in difficult times, and celebrate moments of joy. Emotions play a critical role in shaping interactions, guiding decision-making processes, and influencing overall well-being [6]. In this context, an automated system that keeps family members informed about each other's emotional states can be an invaluable tool.

However, the intersection of facial recognition, emotion detection, and smart home integration presents its share of challenges. As are privacy concerns, detection accuracy and real-time processing are just some of the issues that need to be addressed [7]. This research aims to not only develop a robust facial emotion detection system but also understand its implications, advantages and limitations within the home environment.

The following sections will detail the sections of the project; the literature review, methodology, the development of the alternative, the architecture of the proposed system, the experimental results and the discussions on the impact, the conclusions of the project and the future recommendations of the integration of emotion-aware systems in smart homes.

II. LITERATURE REVIEW

Research in the field of human emotion recognition, especially within contexts like the home, has emerged as a pivotal intersection between technology and psychology. In this section of our study, we delve into the existing literature on this captivating subject. Our goal is to explore prior research conducted by experts in the field, analyzing their approaches, methodologies, and outcomes.

Additionally, we will examine the essential considerations that should be taken into account when addressing emotion recognition in domestic environments, highlighting the opportunities and challenges unique to this particular context. By comprehending the existing knowledge base and drawing lessons from past research endeavors, we will be better equipped to advance our own study and contribute to the growth of this ever-evolving research area.

A. Domestic robot eyes

In the first place a proposed method for recognizing emotions in a domestic environment using a service robot's eyes involves using supervised learning techniques to recognize face and upper-body emotional expressions was found on [8]. The system utilizes two cameras in the robot's eyes model to provide visual feedback for emotion recognition. The cameras should provide images with sufficient resolution for emotion recognition, and the person should be approximately 1m away from the robot's eyes model for better recognition. The system tracks the person in the domestic environment using pan-tilt motion and recognizes the emotional expression for each video frame. The recognized emotion is then displayed on a Graphical User Interface (GUI)

The proposed system for recognizing emotions in a domestic environment [8] using a service robot's eyes has several limitations. For example:

1) Lighting conditions: The system is sensitive to the lighting conditions of the domestic environment. Unwanted silhouettes and variations in lighting can affect the accuracy of emotion recognition . To improve accuracy, the system should be tested and trained under different lighting conditions to make it more robust.

2) Motion blur and occlusion: When the person moves quickly or covers their face with their hands, motion blur and occlusion can occur, leading to errors in emotion recognition . To mitigate this, the system could incorporate techniques to handle motion blur and develop algorithms to handle occlusion, such as using multiple frames or incorporating depth information.

3) Limited training for multiple individuals: The system can currently only be trained and used for one person at a time . To improve accuracy and usability, the system should be extended to recognize emotions for multiple individuals by retraining the system for each person.

4) Limited dataset: The dataset used for training the system may not be comprehensive enough, leading to lower accuracy. To improve accuracy, the dataset should include images with different hairstyles, illumination variations, and a wider range of emotions. This would help the system generalize better to different individuals and situations .

5) Background noise: Large objects with similar colors to the person's skin in the upper-body region of interest can introduce noise and affect emotion recognition . To improve accuracy, the system should be designed to quickly identify and eliminate background noise generated by such objects.

To improve the accuracy of the system, there are some considerations such as:

- Collect a more diverse and comprehensive dataset with variations in lighting, hairstyles, and emotions.
- Incorporate techniques to handle motion blur and occlusion, such as using multiple frames or depth information.
- Develop algorithms to handle background noise and eliminate false detections caused by objects with similar colors to the person's skin.
- Train and test the system under different lighting conditions to make it more robust.
- Explore the use of additional sensors or modalities, such as audio or body movement, to

enhance emotion recognition accuracy.

B. Multimodal Video Recognition System for Caring

The multimodal video recognition system for home caregiving described in the paper employs deep learning models for facial recognition, emotion detection, and pose estimation tasks. These models are trained on extensive datasets, utilizing convolutional neural networks (CNNs) and other deep learning techniques to achieve accurate results. The system runs on a Raspberry Pi, providing a portable and cost-effective solution for real-time video processing and analysis. Additionally, it utilizes the Intel Movidius Neural Compute Stick (NCS) to accelerate computation, although this introduces overheating challenges during extended use. [9]

The system uses deep learning models to recognize and identify faces, facilitating caregiver and patient identification. It also use Deep learning models analyze emotions expressed in the video, offering insights into patient emotional states.

It's important to mention that the Raspberry Pi's limited computing capacity hinders the system, achieving a modest 1.3 FPS when running all models concurrently, with pose estimation being the most computationally intensive. Also, the use of the Intel Movidius Neural Compute Stick leads to overheating problems, particularly during extended operation, impacting both the system and Raspberry Pi.

Likewise, the model selection prioritizes performance on Raspberry Pi but may not deliver the highest accuracy compared to other models, potentially affecting system reliability.

C. Affective Interaction in Domestic Service System

The system proposed in the research is a domestic service system designed to incorporate affective interaction for the evaluation and regulation of emotions in older adults. Its architecture comprises two primary modules: one for evaluating the user's physiological and psychological state based on a behavior model, and another for affective interaction. [4] The principal techniques use on the investigation are :

1) Facial Expression Recognition: Emotions are identified as neutral, positive, or negative through the capture of facial expressions using depth cameras and robot-mounted cameras.

2) User Behavior Model: Daily behavior patterns are studied to establish a behavior model that categorizes activities into entertainment, social activity, and life assistance.

3) Affective Interaction: This technique employs facial expressions and voice interaction to identify emotions, recommend activities, and design personalized intervention scenarios.

4) Emotion Evaluation and Regulation: Emotions are evaluated and regulated using facial expression recognition and the user behavior model, with the design of multimodal affective interactions.

5) Personalized Behavior Model: A behavior model tailored to the user's preferences is created based on caregiver interviews and observations.

6) Multimodal Affective Interaction Design : : The system includes a mobile platform, robotic arm, gripper, and 3D depth camera, facilitating affective interactions through an interface and voice output device.

7) System Architecture: The system architecture encompasses two primary modules: one for evaluating the user's physiological and psychological state based on the user behavior model, and another for affective interaction, including emotion recognition, multimodal interactions, and emotional stimulation.

The study [4] concludes that the system effectively evaluates and regulates the emotions of older adults through affective interaction, including facial expression recognition and a user behavior model. Affective interaction interventions based on this system promote positive emotional states in older adults.

Similarly, the system is acceptable to older adults and fulfills their affective interaction needs [4]. Future research should focus on improving the accuracy of user state evaluation, exploring different behavior model establishment methods, and reliable emotion recognition techniques in dynamic environments. Further user studies are required to enhance the natural interaction experience and validate the system's effectiveness comprehensively.

D. Emotion Recognition from Body Expressions

The research project [1] involves the creation of an emotion recognition system primarily focused on capturing and interpreting body language cues exhibited by individuals expressing six distinct emotions: anger, fear, happiness, neutrality, sadness, and surprise. The data collection involved nineteen participants from various cultural backgrounds, recorded via a humanoid robot Nao equipped with an Asus Xtion depth sensor. Each participant performed each emotion five times, resulting in a dataset of 570 sequences. This depth sensor enabled the system to robustly detect body motion and create 3D skeleton models for further analysis.

The research introduces a neural network architecture specifically designed for recognizing emotions based on body motion patterns within human-robot interaction scenarios. Key methodologies include the utilization of Self-Organizing Neural Networks, encompassing a feedforward network (GM) for pose feature learning, a recurrent network (GP) for motion feature learning, and a recurrent variant of the Growing When Required (GWR) network (GI) for learning prototype sequences and associating symbolic labels with unsupervised visual representations of emotions [1]. The system relies heavily on depth map video sequences captured using the Asus Xtion depth sensor mounted on the Nao robot. These sequences enable robust detection of body motion and facilitate the extraction of 3D skeleton information for further analysis The study underwent several phases, including data preprocessing, where 11 key body joints were selected and post-processed to compute 3D skeleton information. The coordinates of these joints were smoothed by computing medians every 3 frames, resulting in body motion sequences captured at 10 frames per second. The neural network architecture was then trained on 60% of the video sequences and tested on the remaining 40% to assess its classification accuracy, which was empirically optimized. The paper also compared the system's performance to human observers.

In conclusion, the research demonstrated the effectiveness of the proposed neural network architecture in recognizing emotions from body motion patterns, achieving an impressive accuracy rate of 88.8% on the BEE dataset. The findings suggest that body expressions can provide valuable social cues, particularly in scenarios where facial cues are limited or challenging to detect. However, the complexity and variations in body motion patterns across individuals present challenges in achieving perfect recognition. To further enhance the system, future research may explore the integration of audio-visual stimuli, the interplay of body and facial cues, and the development of multi-modal emotion recognition scenarios. Additionally, it would be beneficial to consider a larger and more diverse dataset and address potential cultural variations in expressing emotions.

E. Real Time Testing

Based on the author [3] says that first of all, human faces were detected with Haar Cascade library within 30 images per second of the computer camera. After that, the detected images were sent to the model and the classes they belong to were queried. As a result of the predictions, the possibility of belonging to which class the facial expression was shown on a separate screen and the emotion in which class was higher was overwritten on the Haar Cascade frame. This process was performed on every 30 frames that ocurred every second of the camera image obtained in real time.

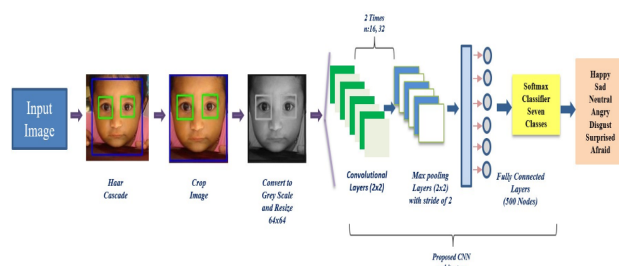

Fig. 1. Proposed Convulutional Neural Network for facial emotion recognition. From: [3]

F. Disturbance Disentangling

The importance facial expression recognition to predict expression by disentangling the disturbance caused by various disturbing factors, such as pose, identity, illumination. [5] propose a different feature learning method to tackle the disturbance caused by facial identity and pose variations. Disturbance-Disentangled Learning (DDL) method to simultaneously disentangle multiple disturbing factors. The method above depend largely on the label information of disturbing factors. Methods address the oclussion problem of Facial Expression Recognition (FER), Region Attention Network to adaptively adjust the importance of facial regions to mitigate the problems of oclussion and variant poses for FER. Some methods are concerned with the noisy label problem in FER databases. Its proposed an Inconsistent Pseudo Annotations to Latent Truth method to deal with the problem of inconsistency in different FER databases. Self-Cured Network to prevent the trained model from over-fitting uncertain facial images.

III. METHODOLOGY

A. Introduction

The aim of this methodology is to provide a step-by-step guide for implementing an emotion detection system in a residential environment using cameras and a Telegram bot. This system will allow homeowners to monitor and receive information about emotions in their house remotely, as [2] stated, some steps mentioned in his document will be used for a better performance . The methodology will cover the following steps:

B. Define Public Spaces for Camera Placement

To ensure effective emotion detection, it is essential to identify the public spaces within and around the house where cameras will be placed. These areas may include the living room, kitchen, entryways, and other common areas where people gather. The placement should prioritize wide coverage and unobstructed views of individuals' faces.

C. Camera Hardware Setup

- Acquire suitable cameras with the necessary specifications, such as resolution and field of view, to capture clear facial expressions.
- Install the cameras at predefined locations within the identified public spaces.
- Ensure a stable power source and internet connectivity for each camera.

D. Software Development for Face and Emotion Detection

- Choose an appropriate software development platform and programming language for the emotion detection system.
- Develop or utilize pre-existing machine learning models for face detection and emotion recognition. Popular frameworks like OpenCV and TensorFlow can be employed.
- Train the emotion recognition model using labeled datasets, if necessary.
- Implement the software to analyze camera feed in real-time, detect faces, and recognize emotions from facial expressions.
- Fine-tune the system to achieve accurate emotion detection.

E. Data Storage and Processing

- Set up a database to store emotion data along with timestamps and camera locations.
- Implement data processing algorithms to aggregate and analyze emotion data for various time intervals (e.g., hourly, daily, weekly).

F. Telegram Bot Integration

- Develop a Telegram bot using Telegram's Bot API.
- Integrate the bot with the emotion detection system.
- Enable secure communication between the system and the Telegram bot.

G. Data Reporting and Notification

- Configure the Telegram bot to receive emotion data and notifications from the system.
- Design a user-friendly interface for homeowners to interact with the bot.
- Implement features that allow users to request real-time emotion updates or historical data.

H. User Authentication and Security

- Implement user authentication mechanisms to ensure that only authorized individuals can access emotion data.

- Apply encryption protocols to protect data transmission and storage.
- Regularly update and patch the software to address security vulnerabilities.

I. Testing and Validation

- Conduct extensive testing to verify the accuracy and reliability of emotion detection.
- Perform user testing to ensure the Telegram bot's usability and effectiveness.
- Address any issues or bugs that arise during testing.

J. Deployment and Monitoring

- Deploy the emotion detection system in the residential environment.
- Continuously monitor the system's performance and reliability.
- Provide ongoing maintenance and support to ensure its proper functioning.

By following these steps, homeowners can create an emotion detection system that allows them to remotely monitor emotions within their house, enhancing security, and facilitating a better understanding of the emotional dynamics in their living space.

IV. ALTERNATIVE'S DEVELOPMENT

Cameras were chosen to be positioned in strategic locations, such as the hallway, living room, and kitchen. These were the places we believed would be most frequented by household members. They were also chosen for these locations because they generally have good light exposure and a good angle to capture faces correctly.

Once the camera locations are finalized, this information will be fed into a cloud solution like Google Vision, which provides us with information about the emotions detected in the captured photos.

With the results obtained from the API, information is sent to the Telegram bot, which will notify the household member of the results.

The importance of facial expression recognition to predict facial expressions caused by various factors such as pose, identity, lighting, etc. proposes several methods to adjust images to mitigate occlusion problems and pose variants for facial expression recognition.

V. RESULTS

VI. DISCUSSION

VII. CONCLUTIONS

VIII. RECOMMENDATIONS

REFERENCES

[1] Nourhan Elfaramawy, Pablo Barros, German I. Parisi, and Stefan Wermter. Emotion recognition from body expressions with a neural network architecture. In *Proceedings of the 5th International Conference on Human Agent Interaction*, HAI '17, page 143–149, New York, NY, USA, 2017. Association for Computing Machinery.

[2] Deepali Joshi, Priti Patil, Vishal Singh, Atharv Vanjari, Tejas Shinde, and Harshada Giri. Face recognition based attendance system. In *2023 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)*, pages 1–6, 2023.

[3] Mehmet Akif Ozdemir, Berkay Elagoz, Aysegul Alaybeyoglu, Reza Sadighzadeh, and Aydin Akan. Real time emotion recognition from facial expressions using cnn architecture. In *2019 medical technologies congress (tiptekno)*, pages 1–4. IEEE, 2019.

[4] Chaolong Qin, Aiguo Song, Yu Zhao, and Mingyang Yin. Affective interaction in domestic service system: Emotion evaluation and regulation for the older adults. In *2022 10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pages 1–4, 2022.

[5] Delian Ruan, Yan Yan, Shenqi Lai, Zhenhua Chai, Chunhua Shen, and Hanzi Wang. Feature decomposition and reconstruction learning for effective facial expression recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7660–7669, 2021.

[6] Ketan Sarvakar, R. Senkamalavalli, S. Raghavendra, J. Santosh Kumar, R. Manjunath, and Sushma Jaiswal. Facial emotion recognition using convolutional neural networks. *Materials Today: Proceedings*, 80:3560–3564, 2023. SI:5 NANO 2021.

[7] Savath Saypadith and Supavadee Aramvith. Real-time multiple face recognition using deep learning on embedded gpu system. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1318–1324, 2018.

[8] T. M. W. Vithanawasam and B. G. D. A. Madhusanka. Face and upper-body emotion recognition using service robot's eyes in a domestic environment. In *2019 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, pages 44–50, 2019.

[9] Chao Yan, Jiahua Xu, Bastian Klopfer, and Andreas Nürnberger. A deep neural network based multimodal video recognition system for caring. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, pages 1–5, 2020.