Identifier of human emotions based on convolutional neural network for assistant robot

Fredy Martínez, César Hernández, Angélica Rendón
Facultad Tecnológica, Universidad Distrital Francisco José de Caldas, Colombia

Article Info

Article history:

Received Sep 2, 2019
Revised Jan 4, 2020
Accepted Feb 23, 2020

Keywords:

Autonomous robot
Convolutional neuronal network
Human emotions
Service robot

ABSTRACT

This paper proposes a solution for the problem of continuous prediction in real-time of the emotional state of a human user from the identification of characteristics in facial expressions. In robots whose main task is the care of people (children, sick or elderly people) is important to maintain a close relationship man-machine, anld a rapid response of the robot to the actions of the person under care. We propose to increase the level of intimacy of the robot, and its response to specific situations of the user, identifying in real time the emotion reflected by the person's face. This solution is integrated with algorithms of the research group related to the tracking of people for use on an assistant robot. The strategy used involves two stages of processing, the first involves the detection of faces using HOG and linear SVM, while the second identifies the emotion in the face using a CNN. The strategy was completely tested in the laboratory on our robotic platform, demonstrating high performance with low resource consumption. Through various controlled laboratory tests with different people, which forced a certain emotion on their faces, the scheme was able to identify the emotions with a success rate of 92%.

Corresponding Author:

Fredy Martínez,
Facultad Tecnológica,
Universidad Distrital Francisco José de Caldas,
Bogotá, Colombia.
Email: fhmartinezs@udistrital.edu.co

1. INTRODUCTION

Service robots are designed to develop tasks that support the human being [1-3]. The success in the development of these tasks is intimately linked to their ability to interact with the user [4-6]. Most of the key to human interaction lies in the ability of one person to feel sympathy for the other, that is, the ability to identify with another person, know how to listen, understand their problems and emotions [7, 8]. Much of this empathy is visualized by observing a person's face. The human face is designed to express a large number of emotions automatically from the emotional state of the individual.

While it is true that the face is not the only means used by humans to communicate their emotional state, the importance of the optical sensor for interaction with the environment makes this a fundamental field to study [9, 10]. However, there are schemes developed to identify other mechanisms for communicating emotions, even in animals. Some of these media include vocal communication [11, 12], body expression [13], forms of direct contact [14] and neurological patterns [15, 16]. Facial expressions on the human face can be used to identify a person's emotional state [17]. That is why many strategies have been developed to identify them from the processing of the image of the face [18-20]. These expressions have common characteristics for a given emotional state, characteristics that are independent of race or gender. If it is possible to identify these images from the analysis of an image, then it is possible to inform to an artificial system, as it is the case of an assistant robot, how to adjust its behavior according to the emotional state of the user.

The most robust solutions developed for the automatic detection of facial expressions use large datasets for the training of adjustable schemes [21-23]. While these schemes provide high-performance values, increasing their reliability with low computational cost has become a challenge due to the wide variability of characteristics between individuals, which makes the separation between classes or categories complex, even without considering the variables related to image capture. The work presented in this paper aims to contribute to this area of research, particularly evaluating the performance of an automatic identification system of emotions in real-time on a laboratory robot assembled to perform tasks of assistance to human beings. In previous works, we have shown details of the robotic platform and a people tracking scheme [24], this work is integrated with such schemes to produce an integrated man-machine interaction solution.

Our robotic platform consists of two integrated robots to form a single healthcare system. We use a Nao robot from SoftBank Group at the top for direct interaction with people, and we use our TurtleBot 1 robot at the bottom to solve navigation and path planning problems. As applied research the research group has studied the problem of safe grasping of cylindrical objects with anthropomorphic hands [25], considering also the delicate interaction with humans. We have also developed many autonomous navigation schemes for locally observable dynamic environments [26-29], and analyzed the problem of dynamic bipedal walking [30]. All these are current unsolved problems of healthcare robotics, which seek to be integrated into the development of more complex tasks such as the monitoring of vital signs in patients, administration of food and medicine or immediate attention in case of emergencies.

The rest of this paper is organized as follows. Section 2 provides some theoretical details of the problem being addressed. Section 3 shows the methodological strategy developed as well as details of the laboratory implementation. This is followed by the evaluation of the system. Finally, we conclude by discussing the lessons learned and how the system could be further developed.

2. PROBLEM FORMULATION

The goal of this research is to develop a robust and high-performance software tool that allows the development of autonomous tasks of identifying human emotions by a small autonomous robot. The work is strongly motivated by the need for this feature as part of the routine interaction of an assistive robot that operates in unknown indoor environments. This sort of robot must navigate an unknown environment, but the main motivation of its action is strongly determined by its level of interaction with the person for whom it works. Let $W \subset \mathbb{R}^2$ be the closure of a contractible open set in the plane that has a connected open interior with obstacles that represent inaccessible regions. Let $\boldsymbol{\mathcal{O}}$ be a set of obstacles, in which each $\mathrm{O} \subset \boldsymbol{\mathcal{O}}$ is closed with a connected piecewise-analytic boundary that is finite in length. The position of obstacles in the environment changes over time in an unknown way, but they are detectable by distance sensors. In addition, the obstacles in $\boldsymbol{\mathcal{O}}$ are pairwise-disjoint and countably finite in number.

Let $E \subset W$ be the free space in the environment, which is the open subset of W with the obstacles removed. This space can be freely navigated by the robot, but it can also be occupied at any time by an obstacle. The robot knows the free space in the environment E from observations, using sensors. These observations allow him to build an information space I. An information mapping is of the form:

$$q: E \to S \tag{1}$$

where S denote an observation space, constructed from sensor readings over time, i.e., through an observation history of the form:

$$\tilde{o}: [0, t] \to S \tag{2}$$

The interpretation of this information space, i.e., $I \times S \to I$, is that which allows the robot to make decisions. The problem can be defined as the definition of a specific filter to apply to the data sensed by the robot's camera in real-time to identify emotional states in the person with whom it interacts. This information must modify the behavior of the robot, both interaction, and movement in the environment.

3. METHODOLOGY

Our scheme of identification of human emotions is composed of two blocks: the first block of face detection, and the second block of identification of emotions in the faces extracted from the video frames as shown in Figure 1. The final result (identified face and emotion) is placed on each frame to reconstruct the video (this is visualized on a 7-inch touch screen on the robot for validation purposes). For face detection we use *dlib.get_frontal_face_detector()*. This face detector is based on the histogram of oriented gradients (HOG) and linear SVM (Support-Vector Machine). There is also a face detector in dlib based on CNN (convolutional neural network), and although it has a better performance detecting faces in many angles (*dlib.get_frontal_face_detector()* only works well with frontal faces), we decided not to use it due to its high consumption of resources, which made it impossible to use in real-time on our assistant robot (the strategy based on CNN was a little more than 15 times slower in each frame than the strategy based on HOG).

The identification of emotions is done through a convolutional neural network that implements the network structure proposed in [31]. This architecture uses global average pooling in its output layer with Sofmax activation function (number of classes equal to the number of feature maps) to eliminate the fully connected layers, and thus considerably reduce the number of adjustable network parameters. The rest of the network is a standard fully-convolutional neural network with a total of nine convolution layers with ReLU (rectified linear unit) activation function to increase sensitivity and avoid saturation, and batch normalization so that each layer of the network learns by itself more features independently of the other layers. The final model contains a little less than 600,000 adjustable parameters and was trained with a dataset of 7,000 images, with 1,000 images in each of its seven categories: angry, disgust, fear, happy, sad, surprise and neutral. The training used 70% of the dataset, and the remaining 30% was used for validation. The accuracy achieved by the model with the validation data was 58%. The code was developed in Python, and we use Keras as the deep learning framework. In Figure 2 we show a sample of the images used for the angry category.

A deep learning framework is an interface, a library or a tool that enables to build deep learning models more easily and quickly, without going into the details of the underlying algorithms. They provide a clear and concise way to define models using a collection of pre-built and optimized components. Instead of writing hundreds of lines of code, it is possible to use a suitable framework that allows a model of this type to be built quickly. There are many machine learning frameworks already in place, and new frameworks are appearing regularly to address specific niches. We are using Keras as a framework (https://keras.io/) with Tensorflow backend (https://www.tensorflow.org).

This emotion identification program was implemented on our assistant robot to work in parallel with other algorithms of specific purpose, in particular with our robot-human basic interaction scheme supported in visual tracking for assistant robot [24]. Our assistive robot consists of two robotic platforms: A humanoid Nao robot from SoftBank Group for interaction with humans and the environment, and an ARMOS TurtleBot 1 robot from the ARMOS research group for indoor navigation as shown in Figure 3. Communication with the two platforms is via a Wi-Fi connection. The algorithm was developed in Python and was programmed on the ARMOS TurtleBot 1 robot. The video is captured by the camera on the head of the Nao robot, and transmitted by Wi-Fi. Processing is performed by the ARMOS TurtleBot 1 robot in real-time in its control unit (DragonBoard 410c of Arrow Electronics with ARM Cortex-A53 Quad-core up to 1.2 GHz per core and Qualcomm Adreno 306 @ 400MHz). We use in our implementation OpenCV 4.1.0, numpy 1.16.2, Keras 2.2.4, TensorFlow 1.14.0, dlib 19.17.0, and imutils 0.5.2.

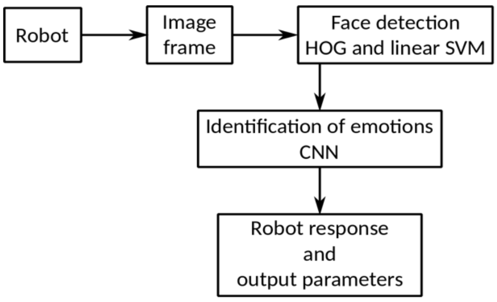
Robot
Image frame
Face detection HOG and linear SVM
Identification of emotions CNN
Robot response and output parameters

Figure 1. Proposed scheme for the identification of human emotions on our robot

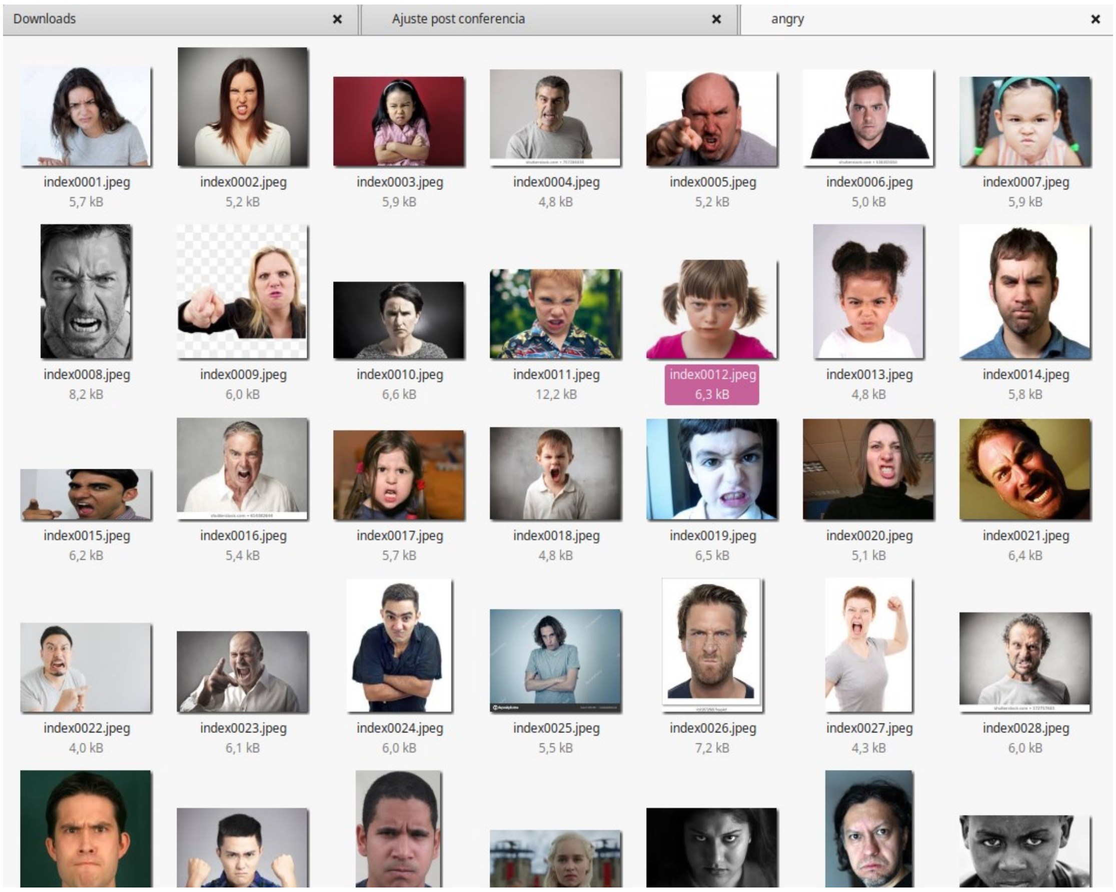

Figure 2. Dataset used for convolutional network training in the angry category

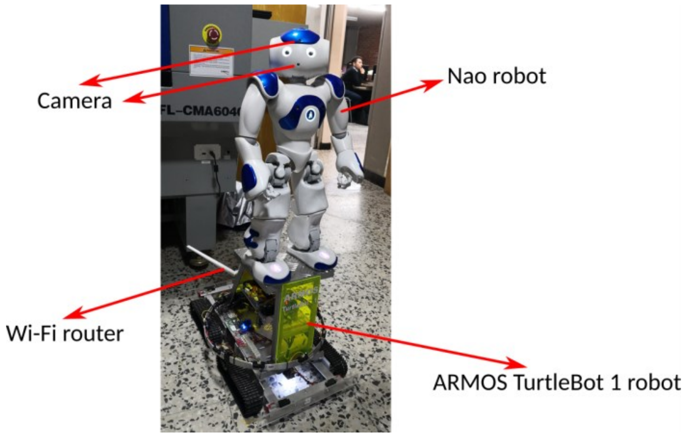

Figure 3. Experimental setup for the emotion identification system. It is composed of a humanoid Nao robot from SoftBank Group at the top and an ARMOS TurtleBot 1 tank robot from the ARMOS research group at the bottom

The images were captured by the Nao robot at a size of 1280✕720 pixels. Each frame was scaled to 632✕480 for processing in the DragonBoard. We conducted several experiments with different lighting conditions, environments, people and distances to the robot. We even performed tests with two or more people at different distances, but always within the robot's field of vision. One of these tests can be seen in: *https://youtu.be/yQajUZTTwmA* as shown in Figure 4.

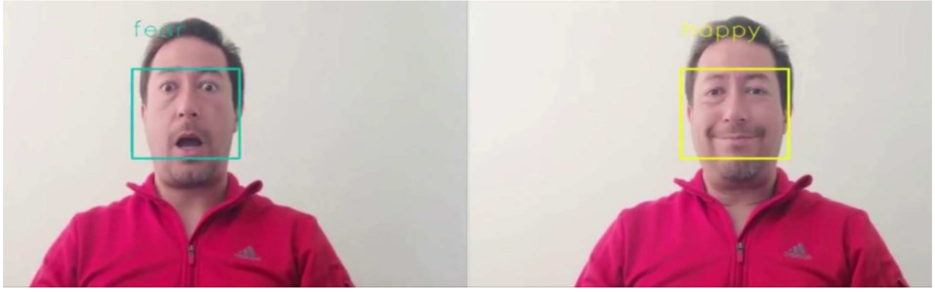

Figure 4. Capture of two emotive states identified during one of the tests developed in the laboratory to the proposed scheme

4. FINDINGS

In general, the robot easily detects people's faces and correctly identifies their emotions. The algorithm demonstrated high performance in laboratory tests allowing real-time operation and parallel execution with tracking and tracing schemes. The flaws in the identification of emotion in a frame were solved by averaging the result of neighboring frames, which resulted in a success rate greater than 92%. As mentioned, the face detection scheme has problems identifying faces that are not completely facing the robot. In the first instance, this has not been a problem for our application since it seeks direct human-machine interaction. However, the results show that it is necessary to evaluate more robust strategies. The emotion identification scheme presents some level of bias concerning the database used for its training. Most of the faces used for training were western people, and this is reflected in the results for some frames. This problem was solved by averaging the behavior along with several frames under the assumption that the person remains with the same emotion a few milliseconds, and therefore is observed in several neighboring frames. It was also observed in some cases problems to identify the emotions when the person has glasses and/or cigarettes, the problem is minor, but also raises the need to improve the training database of the network.

5. CONCLUSION

In this paper, we present the application of a scheme for the identification of human emotions for assistance robots. The objective of this application is to improve the human-machine interaction of our robotic platform. The scheme is structured around two functional blocks, a face detector using HOG and linear SVM, and an emotion identification block using CNN. The output information of the scheme (location of the face and emotion of the person) is used to coordinate the movement and reactions of the robot. The implemented scheme demonstrated a low computational cost, allowing its use in real-time and in parallel with other routines of active tracking of people, this is the most outstanding result of our implementation. The performance evaluation allowed to identify possible improvements to the strategy, in particular, to improve the training of the convolutional network by improving the initial dataset. After averaging the identification capability over several neighboring frames to solve the noise identification problems in the images, the scheme achieved a 92% success rate operating in real time on the robot control unit.

ACKNOWLEDGMENTS

This work was supported by the Universidad Distrital Francisco José de Caldas, in part through CIDC, and partly by the Facultad Tecnológica. The views expressed in this paper are not necessarily endorsed by District University. The authors thank the research group ARMOS for the evaluation carried out on prototypes of ideas and strategies.

REFERENCES

[1] M. Hao, et al., "Proposal of initiative service model for service robot," *CAAI Transactions on Intelligence Technology*, vol. 2, no. 4, pp. 148-153, 2017.

[2] C. Sirithunge, et al., "Proactive robots with the perception of nonverbal human behavior: A review," *IEEE Access*, vol. 7, no. 1, pp. 77308-77327, 2019.

[3] M. A. V. J. Muthugala and A. G. B. P. Jayasekara, "A review of service robots coping with uncertain information in natural language instructions," *IEEE Access*, vol. 6, no. 1, pp. 12913-12928, 2018.

[4] D. P. Losey and M. K. OâMalley, “Enabling robots to infer how end-users teach and learn through human-robot interaction,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1956-1963, 2019.

[5] S. Saunderson and G. Nejat, “It would make me happy if you used my guess: Comparing robot persuasive strategies in social human-robot interaction,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1707-1714, 2019.

[6] L. Chen, et al., “Information-driven multirobot behavior adaptation to emotional intention in human-robot interaction,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 3, pp. 647-658, 2018.

[7] C. Clavel and Z. Callejas, “Sentiment analysis: From opinion mining to human-agent interaction,” *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 74-93, 2016.

[8] U. Jain, et al., “Cubic svm classifier-based feature extraction and emotion detection from speech signals,” *2018 International Conference on Sensor Networks and Signal Processing (SNSP)*, pp. 386-391, 2018.

[9] J. S. Castañeda B. and Y. A. Salguero L., “Adjustment of visual identification algorithm for use in stand-alone robot navigation applications,” *Tekhnê*, vol. 14, no. 1, pp. 73-86, 2017.

[10] L. Cao, et al., “Robust pca for face recognition with occlusion using symmetry information,” *IEEE 16th International Conference on Networking, Sensing and Control (ICNSC 2019)*, pp. 323-328, 2019.

[11] S. A. Reddy, et al., “The decisive emotion identifier?” *2011 3rd International Conference on Electronics Computer Technology*, vol. 2, pp. 28-32, 2011.

[12] M. Sidorov, et al., “Emotions are a personal thing: Towards speaker-adaptive emotion recognition,” *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, pp. 4803-4807, 2014.

[13] A. Kleinsmith and N. Bianchi-Berthouze, “Affective body expression perception and recognition: A survey,” *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15-33, 2013.

[14] Gao Y., et al., “What does touch tell us about emotions in touchscreen-based gameplay?” *Transactions on Computer-Human Interaction*, vol. 19, no. 4, pp. 1-19, Dec 2012.

[15] C. Qing, et al., “Interpretable emotion recognition using eeg signals,” *IEEE Access*, vol. 7, no. 1, pp. 94160-94170, 2019.

[16] R. M. Mehmood, et al., “Optimal feature selection and deep learning ensembles method for emotion recognition from human brain eeg sensors,” *IEEE Access*, vol. 5, no. 1, pp. 14797-14806, 2017.

[17] H. Wang and J. Gu, “The applications of facial expression recognition in human-computer interaction,” *IEEE International Conference on Advanced Manufacturing (ICAM 2018)*, pp. 288-291, 2018.

[18] J. Deng, et al., “Cgan based facial expression recognition for human-robot interaction,” *IEEE Access*, vol. 7, no. 1, pp. 9848-9859, 2019.

[19] P. Tzirakis, et al., “End-to-end multimodal emotion recognition using deep neural networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1301-1309, 2017.

[20] R. Yang, et al., “Intelligent mirror system based on facial expression recognition and color emotion adaptation â â imirror,” *37th Chinese Control Conference (CCC 2018)*, pp. 3227-3232, 2018.

[21] L. Chen, et al., “Three-layer weighted fuzzy support vector regression for emotional intention understanding in human-robot interaction,” *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 5, pp. 2524-2538, 2018.

[22] Z. Liu, et al., “A facial expression emotion recognition-based human-robot interaction system,” *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 668-676, 2017.

[23] H. Meng, et al., “Time-delay neural network for continuous emotional dimension prediction from facial expression sequences,” *IEEE Transactions on Cybernetics*, vol. 46, no. 4, pp. 916-929, 2016.

[24] F. Martínez, et al., “Robot-human basic interaction scheme supported in visual tracking for assistance robot using differential filter and k-means clustering,” *Tecnura*, vol. 24, no. 63, pp. 1-19, 2020.

[25] E. Rodríguez, et al., “Fuzzy Control for Cylindrical Grasp in Anthropomorphic Hand,” *Contemporary Engineering Sciences*, vol. 10, no. 30, pp. 1485-1492, 2017.

[26] F. Martínez, et al., “A Study on Machine Learning Models for Convergence Time Predictions in Reactive Navigation Strategies,” *Contemporary Engineering Sciences*, vol. 10, no. 25, pp. 1223-1232, 2017.

[27] F. Martínez, et al., “Visual identification and similarity measures used for on-line motion planning of autonomous robots in unknown environments,” *Eighth International Conference on Graphic and Image Processing (ICGIP 2016)*, vol. 10225, pp. 1-6, 2017.

[28] F. Martínez, et al., “A Data-Driven Path Planner for Small Autonomous Robots Using Deep Regression Models,” *International Conference on Data Mining and Big Data (DMBD 2018)*, pp. 596-603, 2018.

[29] F. Martínez, et al., “An Algorithm Based on the Bacterial Swarm and Its Application in Autonomous Navigation Problems,” *International Conference on Swarm Intelligence (ICSI 2018)*, pp. 304-313, 2018.

[30] J. Gordillo and F. Martínez, “Fuzzy Control for Bipedal Robot Considering Energy Balance,” *Contemporary Engineering Sciences*, vol. 11, no. 39, pp. 1945-1952, Jan 2018.

[31] O. Arriaga, et al., “Real-time convolutional neural networks for emotion and gender classification,” arXiv: 1710.07557, arXiv.org, 2017.