Nonverbal Communication Assistance System Based on Body Recognition in the Field of Education

Javier Fdez and Michita Imai

Abstract—**Throughout this paper, a system able to provide support for the teacher on the nonverbal communication state of each one of the students is described. This system detects and classifies the facial expression and eye gaze of each student, generating a nonverbal state for each one of them. By using this information, the system is able to determine the classroom environment. It is also able to give suggestions to the teacher in the class so that his/her performance can improve. Furthermore, along with this publication, the different implementations that have been carried out and the first results obtained are described. Lastly, this paper discusses the performance of the system and any future work that may be carried out.**

Index Terms—**Body recognition, education, eye gaze, facial expression, nonverbal communication, robotics.**

I. INTRODUCTION

Nonverbal communication is a key component of almost every field in which human interaction takes place. For example, being able to recognize the gestures and expressions of the other person may help us to express our thoughts or beliefs easily. The existence of this kind of communication has been well known throughout history, although its origin is attributed to Charles Darwin in 1982 [1]. According to Darwin, humans continue to make facial expressions because they have acquired communicative value throughout evolutionary history. In other words, humans utilize facial and body expressions as external evidence of their internal state. In [2] it is mentioned how in just one-tenth of a second it is possible to judge and make a first impression out of a situation by taking into account nonverbal cues.

Therefore, knowledge about this kind of communication is something that people wants to acquire as well as recognize due to the benefits that it provides when expressing and detecting emotions on a daily basis.

Due to the range of applications in which nonverbal communication is presented, the focus of this project will be in the field of education. The publications [3] or [4] are just two examples that show how important it is to study nonverbal communication in this field.

A. Objectives

Since the target problem is the loss of nonverbal communication in education, the main objective is to develop a system able to improve the lecture by providing nonverbal communication assistance to the teacher.

This main objective comprises of four parts:
1) Describe the approach of how to study the nonverbal communication of students.
2) Develop a system able to use that information to give suggestions to the teacher.
3) Implement this system in different educational situations.
4) Discuss the benefits and drawbacks and propose future work to be implemented.

B. Related Works

Despite the fact that there are plethora studies related to nonverbal communication about education, the state of art of this paper focuses on particular scenarios.

Firstly, while it would be possible to study the nonverbal communication that the students can receive from the teacher, this research focuses on information that the teacher may acquire from the students.

Secondly, the implementation of this system focuses on two different situations. In the first one, the teacher is physically in the class whereas, in the second, he/she is not and the class is given by a robot teleoperated by the teacher. However, it is important to highlight the fact that in both cases the class is given by the teacher.

A third possible implementation is the distance learning telepresence. However, since the concept is similar to the first one explained above, it will not be mentioned explicitly but all the contributions mentioned can be also applied to this third one.

In [5], an ANN-based psychological profiling system that monitors multiple channels of facial cues to detect truthful and deceptive human behavioral patterns can potentially be adapted to other human states in different environments. In [6], a humanistic computer teaching system is presented. The core of this system is the effective interaction between teacher and student by creating an emotion-learning model based on theories in Psychology.

Nevertheless, this paper presents new contributions not considered in other papers. These contributions are presented in the next subchapter.

C. Main Contributions

This paper encompasses different ideas from previous works, but also proposes several novel contributions that distinguish this paper from previous ones. These main contributions are:
- Real-time nonverbal communication assistance for educational approaches.
- Classroom environment categorization depending on body recognition.

- Body recognition when launching two cameras simultaneously.
- Discussion about the results obtained when implementing the system into two experiments.

II. SYSTEM OVERVIEW

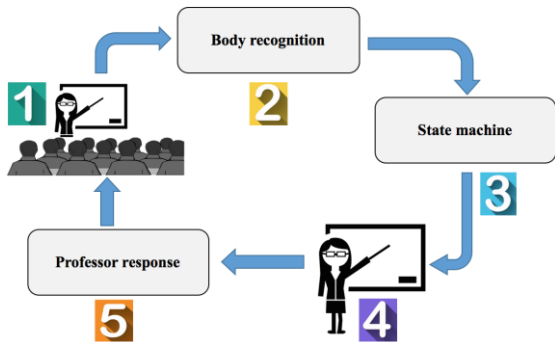Fig. 1 illustrates an overview of the process used for this research.

Fig. 1. General scheme of the system.

First of all, while the class is being given, images of the class are arriving at the system which then classifies the nonverbal communication status of each one of the students that are detected in real time. This classification is done by applying students' body recognition. All this information results in a state machine that gives a classroom environment. Finally, depending on the classroom environment, the system gives suggestions through the interface to the teacher, who is able to respond as he/she deems fit.

The scheme represented in Fig. 1 is shown in more detail in Fig. 2 and Fig. 3.
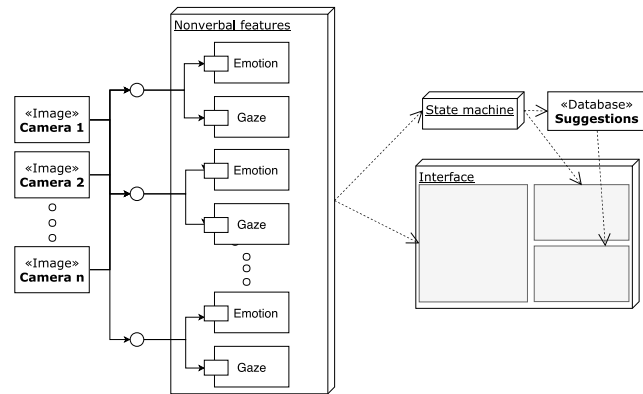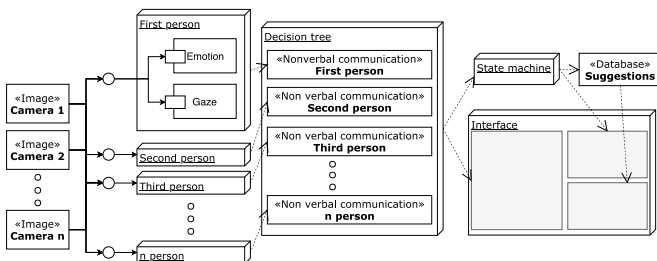

Fig. 2. First logic of the system.

Fig. 3. Second logic of the system.

Each one of the modules of the figures is explained along the following subchapters, as well as the differences on the logic of both figures.

A. Environmental Capture

The human face is not always observed from a single camera. Due to the variety of poses, it is difficult to identify the nonverbal emotion of a person with just one image. Hence, in this project, the detection of environmental information by receiving not only one image but several ones from different angles has been considered.

The outputs of each one of the cameras are the faces detected. These faces are detected by using the frontal face Haar feature-based cascade classifier [7]. This classifier considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and, finally, calculates the difference between these sums. This difference is then used to categorize subsections of an image. The key advantage of a Haar feature over most other features is its calculation speed. Due to the use of integral images, a Haar feature of any size can be calculated in constant time (approximately 60 microprocessor instructions for a 2-rectangle feature) [8].

Once the limitations of the bounding box of the face are detected, the last step is to merge the results obtained from the images of the different cameras, as it is necessary to check what faces are detected by both cameras.

When two cameras view a 3D scene from two distinct positions, there are a number of geometric relations between the 3D points and their projections onto the 2D images that lead to constraints between the image points.

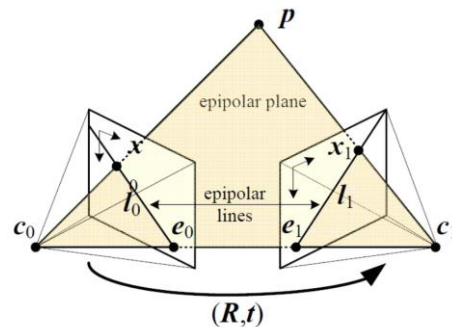In Fig. 4 shows the corresponding lines and plane for one specific point and a pair of cameras.

Fig. 4. Corresponding set of epipolar lines and their epipolar plane.

Consequently, it is possible to find the faces detected by both cameras and merge the results by comparing the bounding boxes, using epipolar geometry as it is shown in the figure above.

In recent years, the perception of eye gaze direction has emerged as a critical component of face processing [9], particularly in tasks requiring visual attention [10]. Therefore, apart from the facial expressions, for this project, the second nonverbal feature that has been considered is the eye gaze of the students.

B. Design of the Emotional Model

Allocating higher precedence to the face than any of the other available channels is regarded as "facial primacy" [11]. Approximately 55% of the communicated message by a person can be obtained from facial behavioral expressions [12] thus making the face a prime area of interest when monitoring human states. Mehrabian [13] argued that even when facial behavior is not consistent with speech, the

listener will be most impacted by the facial emotion expressed. Consequently, there is a great weight toward nonverbal communication being a more reliable source than verbal responses.

In addition, nonverbal communication can be both conscious and subconscious causing that culture, gender and, social status may influence the way nonverbal communication is used. Research on the classification of perceived emotions has centered around the debate between two fundamentally distinct viewpoints. One side of the debate posits that emotions are separate and discrete entities, whereas the other side suggests that emotions can be classified as values on the dimensions of valence (positive versus negative) and arousal (calm/soothing versus exciting/agitating). Paul Ekman, an American psychologist, supported the discrete emotion perspective with his groundbreaking work comparing emotion perception and expression between literate and preliterate cultures [14]. Ekman concluded that the ability to produce and perceive emotions is universal and innate and that emotions manifest categorically as basic emotions (anger, disgust, fear, happiness, sadness, contempt, and, surprise). The alternative dimensional view garnered support from psychologist James Russell, who is best known for his contributions toward the circumplex of emotion. Russell described emotions as constructs which lie on the dimensions of valence and arousal and it is the combination of these values which delineate emotion.

Psychologist, Robert Plutchik, sought to reconcile these views and proposed that certain emotions should be considered as "primary emotions" which are grouped either positively or negatively and can then be combined to form more complex emotions, "secondary emotions", such as remorse, guilt, submission and, anticipation. Plutchik created the "Wheel of Emotions" to outline his theory. This wheel of emotions is shown in Fig. 5.
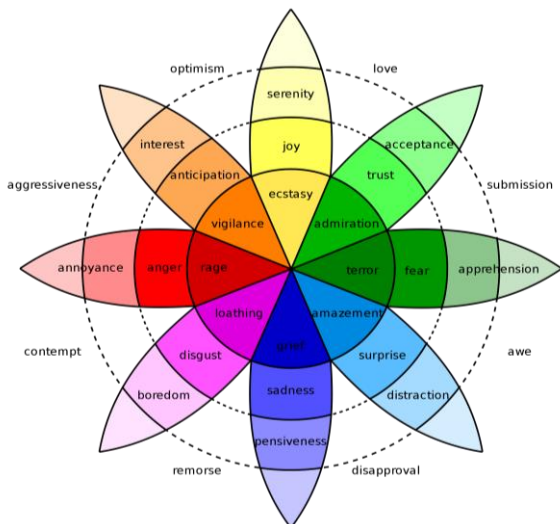

Fig. 5. Plutchik's wheel of emotions.

The primary emotions located on the first ring of the Plutchik's wheel are joy (happiness), sadness, anger, fear, trust, disgust, surprise, and, anticipation, which can have a variety of intensities. Moreover, complex emotions can be obtained from a mix of primary emotions, thus obtaining the full spectrum of human emotions.

For the Nonverbal Communication Assistance System, the emotions considered are six primary emotions (happiness, sadness, anger, fear, disgust, and, surprise) out of the eight explained previously, in addition to neutral emotion (lack of a dominant emotion). Anticipation and trust have been omitted for the Nonverbal Communication Assistance System due to the context of this project.

The detection of the eye gaze when the person keeps the head oriented towards the user results in significant differences between averted and directed gaze, highlighting three emotions: happiness, anger, and, fear [15].

To sum up, this system detects both the facial emotion and the gaze and draws a conclusion about the person's state. The method is shown in Table I:

TABLE I: EMOTION CATEGORIZATION

Facial recognition	Eye gaze	Final state
Happiness	Directed	Happiness
	Averted	Neutral
Sadness	Any	Sad
Anger	Directed	Neutral
	Averted	Anger
Fear	Directed	Fear
	Averted	Fear
Disgust	Any	Disgust
Surprise	Any	Surprise
Neutral	Any	Neutral

C. Students' Nonverbal State Categorization

Once the image arrives at the system, the first step is to detect the facial expression and eye gaze of the students as it has been explained in subchapter II.A. As important it is for the first step of facial recognition to take place, the processing of this information is as crucial.

The state machine, that will be explained in the next subchapter, receives percentages of the emotions presented in the class during a fixed period of time. Consequently, there are two possible approaches.

First, a counter for each one of the emotions will be increased whenever that emotion arrives at the system. After a certain time, the percentage of each emotion, calculated by dividing each emotion by the total number of emotions detected, is obtained. This information is sent to the state machine so that it can determine the classroom environment of the class, as shown in Fig. 2.

The second approach is to apply the same logic but having a counter for each one of the students as shown in Fig. 3. To be able to do this, it is necessary to have a face tracking system. The system follows the next steps.

1) It updates all correlation trackers and removes all trackers that are considered unreliable anymore (e.g. too much movement).
2) For every ten frames, it performs the following:
a. It uses face detection on the current frame and finds all faces.
b. For each face detected, it checks if there exists a tracker for which holds that the center point of the detected face is within the region of the tracker and whether the center point of that same tracker is within the bounding box of the detected face.
c. If such tracker exists, the detected face most likely was already being tracked. If the tracker does not exist, there

is a new face and a new tracker will be created for this face.

3) It uses the region information for all trackers (including the trackers for the new faces created in the previous step) to draw bounding rectangles.

The time set for the system to calculate the percentages, send that information to the state machine and reset the counters is twenty seconds.

It is not a short time, since emotions do not typically change in just few seconds but, it is not also not a long time, which would cause the system not to give many suggestions. Furthermore, a long time period will increase the errors caused by the current system of tracking implemented, specially focused on the logic of Fig. 3.

One important aspect to consider is how the system deals with the circumstance of detecting just the eye gaze or the facial expression, but not both at the same time. For this project, if the face emotion is detected and the eye gaze is not, the student is classified with the output the emotion system is providing. However, in the opposite case, the system will discard the eye categorization and not consider the student's nonverbal state for that particular moment.

D. Classroom Environment Categorization

Regardless of which logic is applied, the input of the state machine is the same and will be the percentage of emotions.

For this project, the categorization of the classroom environment will be simple since the classification will be just neutral, positive or negative states.

Due to the lack of previous researches in which the classification is mentioned, Table II provides the classification used for this project.

TABLE II: CLASSROOM ENVIRONMENT CLASSIFICATION

	Neutral	Positive	Negative
State	Initial state	Happiness > 50%	Rest of the cases
	Neutral > 50%	50%>Happiness> 40% 50%>Surprise > 10%	
	50%>Neutral>40% 40%>Happiness >10%	50%>Happiness > 40% 50%> Neutral >10%	

E. Teacher's Interface

The teacher will have an interface where he/she will first be able to see an image of the class in real time with the corresponding emotions detected. Also, updated current state of the state machine is displayed. Finally, the system displays the three emotions with a higher percentage and one suggestion depending on the current state.

Both the state machine and the suggestions are updated every twenty seconds.

III. METHOD

The Nonverbal Communication Assistance System has been developed with the meta-system operating Robot Operating System (ROS) [16]. ROS provides services designed for a heterogeneous computer cluster such as hardware abstraction, low-level device control, implementation of commonly used functionality, message-passing between processes, and package management. Even if ROS cannot provide real-time

guarantees [17], it is an open-source tool and it allows for a distributed and modular system. Therefore, it is a suitable tool for this project.

For the facial emotion recognition, a Convolutional Neural Network (CNN) building framework has been implemented. The architecture proposed in [18] is used. It proposes a fully-convolutional neural network that contains 4 residual depth-wise separable convolutions. Each convolution is followed by a batch normalization operation and a ReLU activation function. The last layer applies a global average pooling and a softmax activation function to produce the prediction. The input for this neural network are the faces of the people detected, making it possible to classify emotions from different people in a short amount of time. The faces are detected by using the frontal face Haar feature-based cascade classifier. Additionally, the neural network has been trained and tested using the Facial Emotion Recognition-2013 (FER-2013) emotion database. The accuracy of this neural network is 66%.

For eye center localization, the algorithm explained in [19] is used, which efficiently and accurately locates and tracks eye centers in low-resolution images and videos. This last feature and the fact that it is only necessary to detect the eye center in order to classify the eye gaze led into the election of this algorithm. The basis of this algorithm is that, geometrically, the centre of a circular object can be detected by analyzing the vector field of image gradients. The classification of the eye gaze is determined in the following order; first, the Haar feature-based cascade classifier locates the eyes, then the algorithm detects the pupils, and finally, by comparing the position of the eye centre with the edges of the box.

The image sent to the eye gaze classifier is the face detected in the facial emotion classifier so that both emotion and eye gaze are linked to the same person.

The user interface is developed with the software framework RQT, which implements the various GUI tools in the form of plugins. RQT makes it easier to manage all the various windows on the screen at once.

Before closing the methods section, it is explained the methods of the experiments.
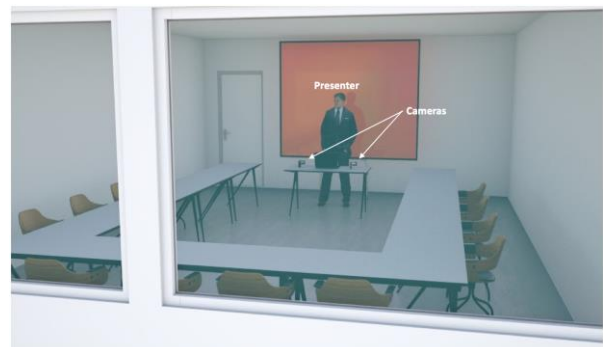

Fig. 6. Recreation of the class.

These experiments were launched at the University of Keio, Tokyo. The experiments consisted of two presentations about different robotic topics in which there was a presenter and ten students. The first presentation was carried out in English and its duration was 4 minutes 52 seconds. The second presentation was presented in Japanese and it lasted 2 minutes 28 seconds.

During these presentations, the classes were recorded with three different cameras. The disposition of the students, the presenter and the cameras is illustrated in Fig. 6.

The disposition of the cameras is, in the case of having two cameras, ten degrees looking to the outside to detect the maximum number of faces. When there is only one camera, it is positioned in front of the computer (this disposition is not shown in the figure above).

IV. RESULTS

The two objectives for implementing the experiments were one, to compare the results between the implementation of just one camera and two cameras simultaneously and two, to analyze and discuss the output of the system when using two cameras taking into account the different logic.

Consequently, this chapter is divided into two subchapters, mentioning each one of them the pertinent results. Further discussion is considered in a future chapter.

A. Results When Different Number and Disposition of Cameras

An image captured from the left camera of the output from the system is shown in Fig. 7.

In this image, it is possible to observe that the faces and eyes detected have a bounding box around them. The emotion for each one of the faces can be also observed and a number is given to each one of the students. The number is determined by the face tracking system.

Fig. 7. Features selected.

The numbers of faces detected in the case of implementing one or two cameras are mentioned in Table III.

TABLE III: CLASSROOM ENVIRONMENT CLASSIFICATION

	English presentation	Japanese presentation
One camera	3047	1057
Two cameras	4291	1439

Further analysis and how new approaches would influence the results for both subsections are discussed in the next section: Discussion.

B. Results of the Emotion and Classroom Environment Classification

Out of each one of the presentations, it is possible to classify the emotions and, consequently, the classroom environment.

As previously mentioned, there are two different approaches when studying the classroom environment: logic of Fig. 2 and Fig. 3. Therefore, this subsection is divided into two sections, one for each one of the logics, displaying the results obtained in each one of them.

In both sections, there are four different graphs. For the first logic, two graph illustrate the percentage of emotions in both presentations for each period of twenty seconds. The other two give the final classroom environment for each period using the criterion explained in Table II.

The second logic also shows final classroom environment. However, since the system bases are focused on the individual state, the latter two graphs display the percentage of each emotion for each period of twenty seconds for just one person.

- Results of implementing Fig. 2 logic.
- Results of implementing Fig. 3 logic.

V. DISCUSSION

Upon analysis and evaluations of the results of the system, the following section discusses the benefits and drawbacks of the current system.

A. Features Selected

The two features selected were eye gaze and facial expression. The performance of these features among different people as well as the performance of the system with another feature are discussed.

As written in [20], only a subset of four emotions out of the six "universal" facial expressions are universally recognized; happiness, surprise, anger and, sadness. This suggests a universal language of fewer than six facial expressions signals. Interestingly, this subset does not include the facial expression signals of fear and, disgust, although both are largely considered to be "primitive" based on the biological origins and association with rapid deep brain activity. Including eye gaze supplements in the recognition of the emotion fear. Therefore, the main problem arises when recognizing the emotion of disgust. Also, when considering the 35, 887 images of the FER2013 database, only 2% of the images portray the emotion of disgust.

Overall, a further effort is needed when detecting and classifying disgust.

Another approach to this project is how the classification of the nonverbal communication has been developed by just considering facial expression and eye gaze. There are more nonverbal features such as the tone of the voice, appearances or physical actions. Among these features, the most relevant one is the study of the body actions which have been carried out in projects like [21].

Even if the evaluation of this feature presents some drawbacks, such as the lack of universal communication among body actions or the difficulty to train the database, which would explain the lack of projects about this field compared with the publications about facial expression, it can be interesting to study in the future and evaluate its performance.

B. Conditions of the Experiments

Some of the conditions in which the experiments were carried out have already been mentioned previously. This subchapter will provide a more detailed explanation of the two determining factors in the results of the experiments.

Firstly, the disposition of the class and the seating arrangements of the students compared with the presenter effect some of the results. Four of the students were facing

the lecturer while the six remaining students were ninety degrees compared to that of the presenter. This increased the difficulty of face detection and classification.

Secondly, the experiment was carried out under conditions in which students were not going to be evaluated. This influenced their interest and motivation to listen to the presenter. Consequently, this could have determined the number of faces that could be detected.

C. Results of the System When Implementing One or Two Cameras

From the results shown in the section before it can also be observed how, no matter which experiment it is developed, the number of faces detected increases. Even if this result could be considered as obvious, there are two considerations to take into account when launching the system.

The first consideration is the position of the cameras. From the results of Section IV.A, it can be observed that the number of faces does not double when another camera is considered. Pointing the cameras in the direction of the students is recommended since it increases the amount of faces detected.

The second consideration is to take into account the height and position of the cameras. In order to achieve optimal accuracy, the cameras should be placed as close as possible from the speaker as well as having the same height as the face of the students. This will ensure the system to detect the frontal view of the faces.
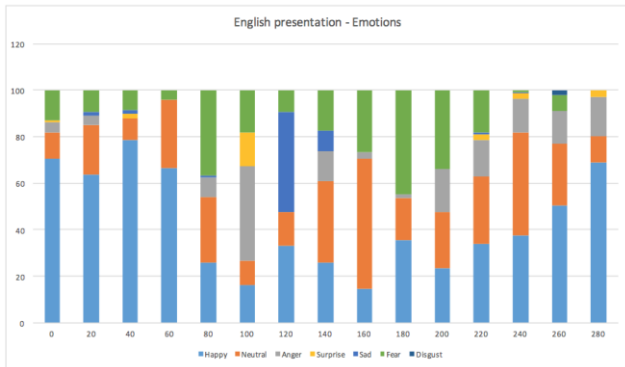

Fig. 8. Percentage of emotions during the English presentation for each period of twenty seconds.
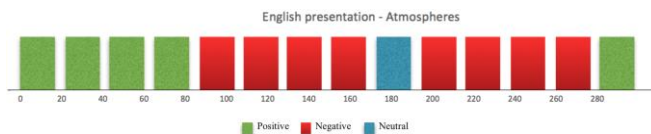

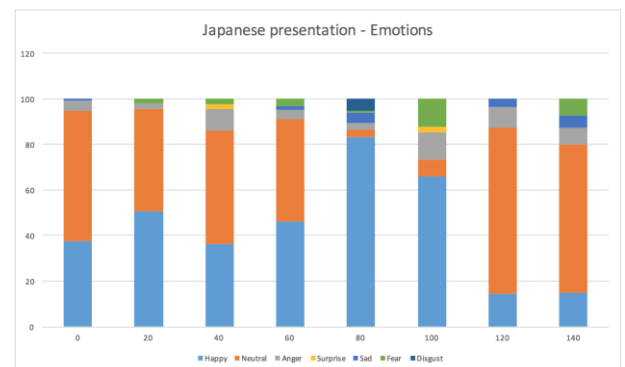Fig. 9. Classroom environment classification during the English presentation for each period of twenty seconds.

Fig. 10. Percentage of emotions during the Japanese presentation for each period of twenty seconds.

D. Results of the System for Positive and Negative Classroom Environment

The different results of the experiments are presented throughout Fig. 8 to Fig. 15. Although the particular results of each one of the experiments could be studied, the focus of this subsection will be to compare and contrast the performances between the logic of Fig. 2 and Fig. 3.
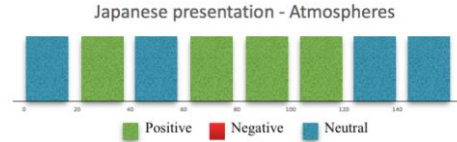

Fig. 11. Classroom environment classification during the Japanese presentation for each period of twenty seconds.
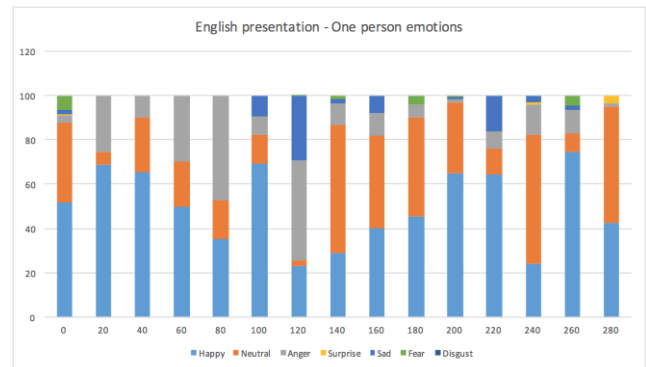

Fig. 12. One person's percentage of emotions during the English presentation for each period of twenty seconds.
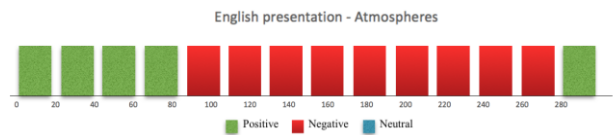

Fig. 13. Classroom environment classification during the English presentation for each period of twenty seconds.

The results between Fig. 9-Fig. 13 and Fig. 11-Fig. 15. presents several differences. In the case of the English presentation, there is one classroom environment that changes whereas in the Japanese presentation, there are two. Despite the differences, this leads to one important conclusion which is that, even if the results of the logic of Fig. 3 should give more accurate results that the logic of Fig. 2, the results of this second logic also gives accurate results.

Therefore, based on the results obtained, depending on the time available to calibrate the cameras and the disposition of the class, the logic implemented should be the one displayed in Table IV.

"Time available" refers to the time to calibrate and measure the position of the cameras. The images received by the system have to detect different people, avoiding the possibility of detecting the same person by both of the cameras simultaneously.

E. Future Work

Throughout the paper, potential uses of the system have been indicated. The following summarizes the possible future usages:

- Develop a subdivision of the positive and negative classroom environment classification and its link with the nonverbal states of the students in the class.

- Input new nonverbal communication features, such as body action, and study the improvement of the system's accuracy.
- Study the correlation between the output obtained from the system and the real perception of the students about the lecture.
- Implement the system in new settings in which there are different dispositions and number of students, a different number of cameras and occlusion between students.

TABLE IV: EXPERIMENT IMPLEMENTATIONS

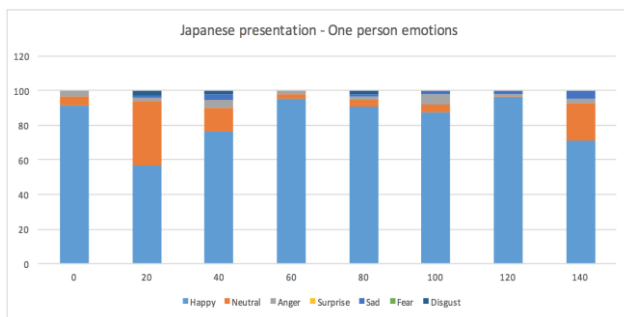| Time available	Disposition of the class	Logic	Cameras
Yes	Frontal	Logic Fig. 2 or Fig. 3	Camera ≥ 1
Yes	Other	Logic Fig. 2 or Fig. 3	Camera ≥ 2
No	Frontal	Logic Fig. 2	Camera ≥ 1
No	Other	Logic Fig. 2	Camera ≥ 2

Fig. 14. One person's percentage of emotions during the Japanese presentation for each period of twenty seconds.
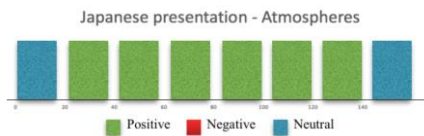

Fig. 15. Classroom environment classification during the Japanese presentation for each period of twenty seconds.

VI. CONCLUSION

This work presents a system able to provide suggestions to the teacher depending on the nonverbal communication state of the students. For this purpose, some experiments were conducted. Moreover, the results of the system are presented and discussed.

Regarding the logic of the system, its configuration is presented with details with this publication, complementing other systems that have already developed.

In conclusion, future works will continue this line of research, trying to improve the system by including more classroom environments, a better categorization and a more solid database that assess the teacher. Furthermore, next efforts will focus on performing more experiments and drawing further conclusions that can be useful for the educational community.

REFERENCES

[1] B. Pease and A. Pease, *The Definitive Book of Body Language*, New York: Bantam Books, 2004.
[2] V. P. Richmond, "Power in the classroom VI: Verbal control strategies, nonverbal immediacy and effective learning," *Journal of Communication Education*, vol. 35, pp. 43-55, 1986.
[3] C. P. Zeki, "The importance of non-verbal communication in classroom management," in *Proc. Conference on Educational Sciences*, 2009, vol. 1, pp. 1443-1449.
[4] John J. Okon, "Role of non-verbal communication in education," *Mediterranean Journal of Social Sciences*, vol. 2, pp. 35-40, 2011.
[5] J. Rothwell, Z. Bandar, J. O'Shea, and D. McLean, "Silent talker: A new computer-based system for the analysis of facial cues to deception," *Applied Cognitive Psychology*, vol. 20, pp. 757-777, 2006.
[6] X. Meng, Z. Wang, G. Wang, L. Shi, and X. Wu, "The research of a teaching assistant system based on artificial psychology," *Affective Computing and Intelligent Interaction*, vol. 3784. 2005.
[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. Conference on Computer Vision and Pattern Recognition*, 2001.
[8] Viola and Jones, "Rapid object detection using a boosted cascade of simple features," *Computer Vision and Pattern Recognition*, 2001.
[9] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161-1178, 1980.
[10] M. Bindemann, A. M. Burton, and S. R. H. Langton, "How do eye gaze and facial expression interact?" *Visual Cognition*, vol. 16, pp. 708-733, 2008.
[11] M. L. Knapp and J. A. Hall, *Nonverbal Communication in Human Interaction*, 3rd ed., Fort Worth: Harcourt Brace, 1992.
[12] A. Mehrabian, "Communication without words," *Psychology Today*, vol. 2, no. 4, pp. 53–56, 1968.
[13] A. Mehrabian, *Silent Messages*, 5th ed., California: Wadsworth Publishing Company, 1971.
[14] P. Ekman, "Facial expression of emotion," *American Psychologist*, vol. 48, pp. 384-392, 1993.
[15] N. Bee, S. Franke, and E. André, "Relations between facial display, eye gaze and head tilt: Dominance perception variations of virtual agents," in *Proc. 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*, pp. 1-13, 2009.
[16] Morgan Quigley, Ken Conley *et al.*, "ROS: an open-source Robot Operating System," in *Proc. ICRA Workshop on Open Source Software*, 2009, vol. 3, pp. 5.
[17] H. Wei, Z. Shao, Z. Huang, R. Chen, and Y. Guan, "RT-ROS: A real-time ROS architecture on multi-core processors," *Future Generation Computer Systems*, 2015.
[18] O. Arriaga, M. Valdenegro-Toro, and P. Plöger, "Real-time convolutional neural networks for emotion and gender classification," *arXiv*.
[19] T. Barth, "Accurate eye centre localization by means of gradients," in *Proc. the Int. Conference on Computer Theory on Computer Theory and Applications (VISAPP)*, 2011, vol. 1, pp. 125-130.
[20] R. E. Jack, "Culture and facial expressions of emotion," *Visual Cognition*, vol. 21, 1248-1286, 2013.
[21] S. Saha, S. Datta, A. Amit, and R. Janarthanan, "A Study on emotion recognition from body gestures using Kinect Sensor," in *Proc. International Conference on Communication and Signal Processing*, 2014, pp. 56-60.

Javier Fdez was born in Madrid, Spain in 1994. In 2016, he received the bachelor's degree in industrial technologies at the Universidad Politécnica de Madrid, Spain, also, he started the master's degree in industrial Technologies at the Universidad Politécnica de Madrid. In 2017, he decided to do the double master's degree program between the Universidad Politécnica de Madrid and the University of Keio, Japan, where he started to study the master of science and technology. Currently, he is working on his master's thesis so that he can graduate in August 2019.

From 2015 to 2017, he received two collaboration grants that allowed him to do research in 3D reconstruction, virtual reality or gait development in a hexapod robot. Out of his research, he published a paper entitled "Integration of reconstruction 3D and virtual reality: A new approach for immersive teleoperation" in the 2017 Third Iberian Robotics Conference.