Feature Extraction from Body Postures: Towards an Emotion Recognition System Based on Digital Imaging

Bruno Barbosa, António J. R. Neves
DETI / IEETA
University of Aveiro
3810-193 Aveiro, Portugal
Email: {brunobarbosa, an}@ua.pt

Sandra C. Soares[1], Isabel D. Dimas[2]
[1]CINTESIS.UA, Departamento de Educação e Psicologia
[2]GOVCOPP/ESTGA
University of Aveiro
3810-193 Aveiro, Portugal
Email: {sandra.soares, idimas}@ua.pt

Abstract — **Recent advances in the study of emotions show that body expressions or body postures are also a relevant non-verbal cue for communicating emotions, with recognition rates similar to those transmitted by facial expressions. Despite the tremendous technological advances in the last decades, there are still room for improvement in the definition of how posture features can be related to different emotions, using technological solutions. This is a complex task due to the fact that it is difficult to associate an emotion with a certain posture and there are a large variety of emotions, such as, sadness, joy, anger, fear, surprise, disgust, among others. In this paper we present a computer vision system that is able to extract characteristics regarding the human body and generate data that allow the evaluation of a given posture. We propose the use of a recently developed algorithm to extract the body posture based on digital color images, which has presented optimal results in several domains, for the recognition of emotions from human body postures. Based on this algorithm, a system was developed that allows, in a non-invasive way, to extract characteristics of the human posture in order to perceive the emotions associated with each posture. We present experimental results in a real scenario, showing promising results regarding the extraction of characteristics obtained from the human body (displacements, areas and overlaps), confirming the effectiveness of the developed system.**

Keywords - Pose Estimation; Digital Image; Emotions; Skeleton Detection.

I. INTRODUCTION

This article is an extended version of the original paper presented at the Third International Conference on Advances in Signal, Image and Video Processing, *SIGNAL 2018* [1]. This version extends the conference paper by presenting a system developed for the extraction of characteristics of the human posture and provides experimental results on a real scenario, confirming the effectiveness of the developed system.

Emotions conveyed by facial expressions are powerful non-verbal cues for functional socio-emotional interactions. The study of body postures as another important non-verbal means to communicate emotions and behavioral intentions has been exponential in the past decade [2], particularly in the fields of cognitive, affective and social neuroscience [3][4]. Although these studies have been showing that emotion recognition performance depicted from body postures do not seem to differ from those of facial expressions, research work exploring the effectiveness of computer vision systems able to automatically detect and classify emotional categories and dimensions from human postures are scant.

Herein, we present the state of the art regarding the development of computer vision systems for detection and classification of human body posture, including the type of existing sensors to obtain images that feed such systems and their operation, as well as human skeletal detection algorithms. We discuss the implications regarding the development of such systems for emotion detection from body posture in several contexts, in which emotions are relevant for socio-communicative purposes.

With the advancement in the study of emotions associated with body postures, it is necessary to investigate, technologically, how emotions can be extracted, non-invasively, from human postures. This is of high relevance to several areas of application, ranging from education, (e.g., posture of students in classrooms, denoting disinterest or excitement [5]), to teamwork [6] and mental health contexts (e.g., postures associated with psychopathology, such as unipolar depression [7]).

Our proposal is that, by using digital cameras and algorithms that allow to extract human body postures, postures associated with different emotional dimensions can be mapped. We present a system to extract the characteristics of the human posture that allows the detection and extraction of characteristics of these postures, in groups of individuals who are asked to freely interact in a dynamic way, thus allowing each posture to be classified according to the associated emotion. Hence, unlike the previous studies, participants will not be asked to perform specific postures that are expected to be associated with different socio-communicative patterns (e.g., expansive or constrictive) [8][9][10]. This raises the following questions: "How to use a PC and a camera to estimate the body posture of the human body?" and "How to classify each posture as being associated with certain emotional dimensions and or categories?".

This article is organized as follows: Section I gives a brief introduction and presentation of problem, in Section II the existing computer systems for the described problem are presented, Section III discusses the definition of digital image, presenting the various types of existing image sensors, in Section IV are addressed some of the existing posture detection algorithms. Section V presents the feature extraction system developed and Section VI shows the results obtained by this system. Finally, in Section VII, a conclusion is drawn based on the results and effectiveness of the system.

II. Existing Computational Systems for the Presented Context

Some work has been carried out in the development of systems for the detection and evaluation of the human body. However, no system was yet developed to allow the detection and classification to map emotions from body postures. This lack of systems' is due to the difficulty of classifying a posture. Moreover, in real life settings, the variations in postures are immense, making it difficult to infer emotions from dynamic interactions between individuals.

In [11], in a classroom context, the authors claim to provide important information to the teacher about their audience's attention. This study focused mainly on the capture of data through a camera system to detect movements, as well as the head and its orientation, thus obtaining the most significant patterns of behavior to infer this cognitive dimension (i.e., attention). However, the results failed to show a direct relationship between the movements of the students and their attention.

In [12], a system for the recognition of human actions based on posture primitives is described. This system, like [13], only focuses on perceiving/classifying if a person runs, walks, dance, etc. and not their emotions. In a learning phase, the representative parameters of posture are estimated through videos. After that, already in a classification phase, the method is used for both videos and static images. In this system, 3 disjoint problems are identified in the recognition of human action: detection of a person in the image, recognition of the posture expressed, and attribution of a category of action to its posture, the focus being the last 2 points. The results of this system are promising, resulting in a highly accurate recognition of actions, allowing us to conclude that the posture of human beings contains enough information about their activity. It is also mentioned that, the addition of other information besides posture, allows for a greater precision in the recognition of the activities.

In short, we were able to verify the existence of some systems for the recognition of posture with specific applications. However, no system is yet available to recognize and classify postures according to the emotions they are communicating.

III. Image Sensors

Typically, a digital image is represented by a rectangular matrix of scalar vectors, composed by a finite number of elements in each position and with a certain value. These elements are called pixels [14].

A pixel is the smallest unit of an image and has an intensity value and a location associated with it. Through the joining of many pixels and due to the filtering effect of the human eye, it is possible to create illusions, like gradients and shading.

Figure 1 represents the gradient of a Red, Green and Blue (RGB) image by merging pixels.

The most common types of digital images are grayscale and RGB images. In grayscale images, the value associated with each pixel is black, white or a shade of gray, which can range, for 8 bits per pixel, from 0 to 255, where 0 is black and 255 is white. In color images, each pixel has associated with it a red, green and blue value, which combined in different amounts can generate any color. The values of red, green and blue also vary, for 8 bits per pixel, between 0 and 255, with 0 being the black color and 255 the maximum of the respective color. Figure 2 shows an intensity matrix of a grayscale image for a given area [16].

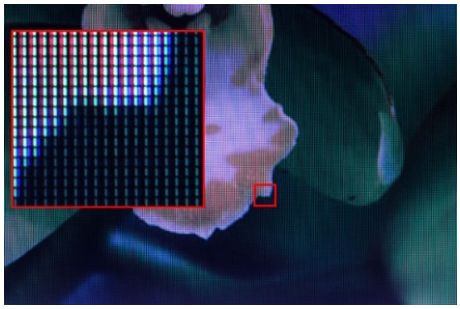

Figure 1. Gradient associated with a region of an RGB image [17].

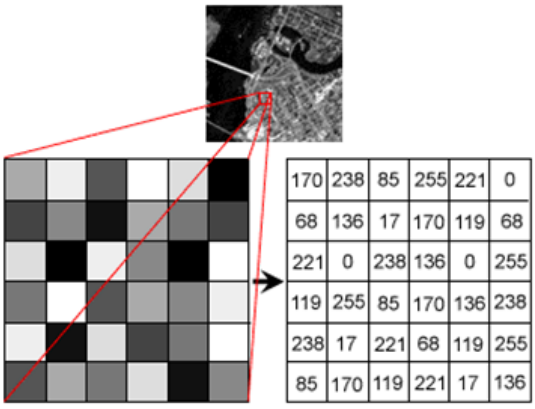

Figure 2. Matrix for a certain area of a grayscale image [18].

The resolution of a digital image depends on the size of its array, that is, with increasing number of pixels, the resolution increases. However, the processing of this matrix becomes computationally slower.

There are several types of sensors able to obtain digital images. In the next subsections, some of these types of sensors will be discussed and their operation will be explained.

A. Image Sensors in the Visible Spectrum

For capturing digital images in the visible spectrum, mainly two types of sensors are used - the Charge-Coupled Device (CCD) and the Complementary Metal-Oxide-Semiconductor (CMOS) sensor.

Each of these sensors is composed by millions of photosensitive transducers whose function is to convert light energy into electric charge. They also have a photosensitive surface, which receives a charge of light to capture the image, so the larger the photosensitive surface, the better the image quality [19].

However, these sensors can only measure the energy of the radiation. To obtain color images, it is necessary to apply a filter that allows to target specific colors to their respective pixels. The most common filter is the Bayer filter. Figure 3 shows the operation of this type of filter.

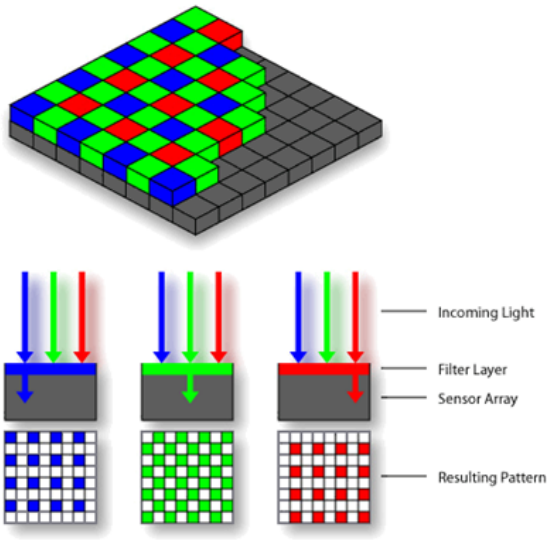

Figure 3. Application of a Bayer filter to obtain a color image [20].

The CCD sensor exists mainly in compact cameras, while the CMOS sensor is present from simple webcams and smartphone cameras to professional cameras.

Figure 4 shows an example of a CCD and CMOS sensor.

Figure 4. Example of CCD (left) and CMOS (right) sensor [21].

B. Special Sensors

In addition to the sensors mentioned earlier, there are also special sensors that allow to obtain other information besides the color image. These sensors are especially used for image processing in special cases, such as the measure of distances and temperatures.

In the next subsections, the modes of operation of these sensors will be explained.

1) Thermal

A thermal camera, unlike the cameras in the visible spectrum mentioned above, are composed of sensors capable of capturing radiation in the infrared spectrum, thus allowing the creation of an infrared image [22]. Normally, when displaying this type of images, a color table is applied so that it is possible to easily distinguish between hot and cold zones. Figure 5 shows a thermal image, obtained through a *Flir* [23] thermal camera, with the respective color table. Although this camera is commercial, it has a high cost due to its specific market and technology used in its manufacture.

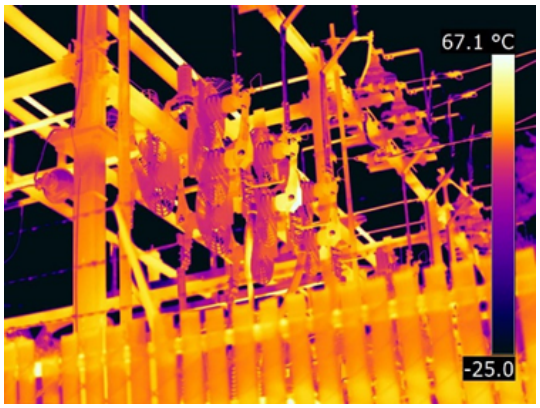

Figure 5. Example of a thermal image, obtained through a *Flir* [23] thermal camera.

This type of sensor can be used even in low-light environments, as opposed to sensors, such as CCD and CMOS [22]. There are several areas where these apply. From security, where they can be used to detect intruders even in low light situations [24], to the industry, where they can be used to detect heating problems in machines, which are not detected by the human eye [22], passing through the detection of people through the temperature of the human body [25].

2) Multi/Hyper Spectral

The Multispectral and Hyperspectral sensors measure the energy in various bands of the electromagnetic spectrum. The spectral resolution is the main distinguishing factor between the images produced by these two types of sensors.

The hyperspectral sensors contain a greater number of bands with narrow wavelengths, providing a continuous measurement in all the electromagnetic spectrum, whereas the multispectral sensors usually contain between 3 and 10 bands with wide wavelengths in each pixel of the image produced [26]. This way, the images captured by a hyperspectral sensor contain more data than the images captured by multispectral sensors. In a practical context, images produced by multispectral sensors can be used, for example, to map forest areas, while images produced by hyperspectral sensors can be used to map tree species within the same forest area [27].

Figure 6 shows the comparison between multispectral and hyperspectral images.

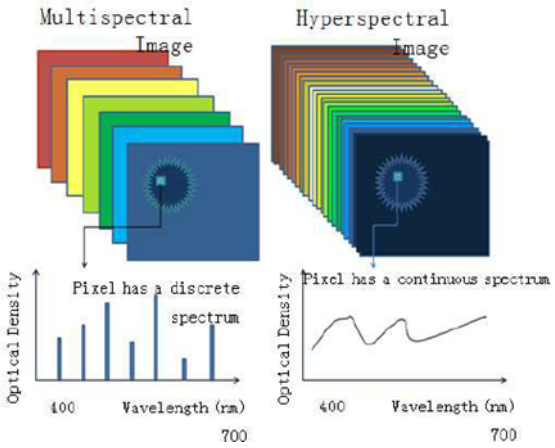

Figure 6. Comparison between a multispectral (left) and hyperspectral (right) image [28].

3) Distance

There are several types of distance image sensors. These types of sensors can obtain images where the closest and most distant objects are perceptible.

There are three major types of sensors, the sensors called Time Of Flight (TOF), Structured Light and Stereo. TOF sensors work on the principle of sending and receiving a signal by measuring the properties of the received signal. By determining the flight time and, consequently, through this time and the speed of the signal the distance to the object is obtained [29]. Structured Light sensors work by projecting a previously established pattern into scene, allowing the system, by capturing that same pattern, to calculate the depth of each pixel of the image received. This calculation is performed by deformation of each point of the pattern projected in combination with the original pattern [30]. Finally, the Stereo sensors allow to obtain distance image through two lenses, at a certain distance, so that the two captured images can be processed and compared, creating a 3D image [31].

IV. Algorithms for Posture Detection

There are many human posture detection algorithms, but few do it dynamically and in poorly controlled environments.

The main existing algorithms focus on the area of vision. This area has been increasingly explored as it allows everything to be done in a non-invasive way for the human being. Thus, devices not directly in contact with it enable the ecological validity of the actions, hence increasing the accuracy and credibility of the algorithm. In this type of algorithm, the detection is done using external objects such as flags [10], or simply through the previous teaching of the system for the intended postures [9].

A posture emerges as well as the set of 2D or 3D locations of the joints, being possible, through these locations, to assess the position and displacement of all limbs. However, the problem that is common to these algorithms relates to critical body positions, such as lying, sitting, shrunken, sideways, etc. [9][32] and in situations that involve groups of people, where some parts of the body overlap [32]. In this type of positioning, the accuracy of these systems drops significantly.

All posture detection algorithms presented here are based on videos or a set of images collected from digital cameras. There are thus several types of cameras used with these algorithms. As described in the previous section, these cameras may differ in the type of image you can get. However, at present, the Kinect is the preferred device of most of these algorithms, since its own Software Development Kit (SDK) is one of the most used with respect to detection of the human skeleton. Kinect consists of an RGB camera, depth sensor, a three-axis accelerometer, a tilt motor and a microphone vector [33]. Thus, it is possible to obtain, with only one device, different types of images. Figure 7 shows the various components of a Kinect.

As mentioned previously, its software, Kinect Skeletal Tracking, is widely used in the detection of the human skeleton, which is carried out in three steps. In the first, an analysis, per pixel, is made to detect and classify body parts; in a second phase, a global centroid is found to define the joints of the body; finally, a mapping of the joints is done, so that they fit into a human skeleton, through data previously known about the human skeleton [34]. Figure 8 shows the steps explained above.

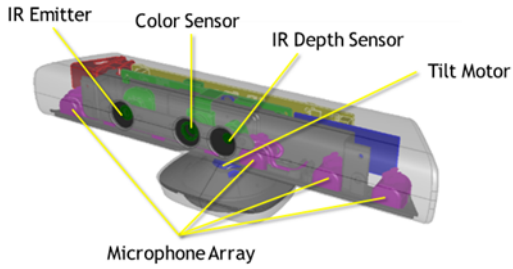

Figure 7. Hardware Configuration of a Kinect Device [35].

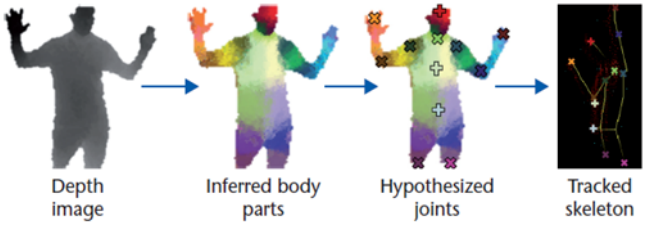

Figure 8. Detection steps of the Human Skeleton through the Kinect Skeletal Tracker Software [34].

In April 2017, the OpenPose library [36] was launched. Using only RGB images, this library can detect and extract 2D values from the main parts of the human body. With this library, it is possible to perform a detection of body, face and hands, in a total of 130 possible keypoints, 15 or 18 of them for body parts, 21 for each hand and 70 for the face.

For body detection, one of two data sets are used: Common Objects in Context (COCO) or MPII Human pose dataset, with people images, annotated with the human skeleton, still being

used CMU Panoptic dataset during the development of the algorithm, since it contains about 65 sequences of approximately 5 hours and 30 minutes and 1.5 million 3D skeletons available. This detection is done through the approach described in [32], where a neural network is used to simultaneously predict confidence maps for body part detection (Figure 9b) and affinity fields for association of parts of the body (see Figure 9c), this process being done in several steps, so that this detection is credible.

Next, a set of two-part combinations is performed to associate the body parts, where a score is used to define which person belongs to the respective part and to make a correct connection of the parts in each person in the image/frame (Figure 9d). Through this approach, it is possible to detect several people in the image and define their posture. Finally, with a greedy inference algorithm, all parts are connected and the 2D points are defined for each of the joints (Figure 9e).

In [37][38], thre are presented approaches of detection multiple human skeletons in simple RGB images with efficient results, however they fall short of [32].

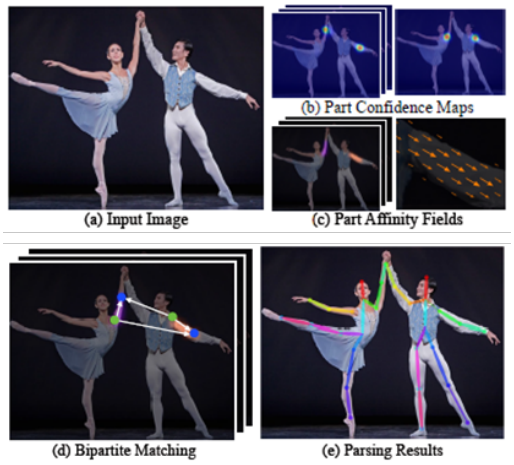

Figure 9. Detection of the Human Skeleton through the OpenPose library [32].

V. System for Features Extraction of Human Posture

In this paper we propose a system to extract the characteristics of the human posture contributing to the presented technological gap for the presented context. The system was developed based on the use of the OpenPose algorithm and RGB images to detect the human body posture and extract characteristics necessary for its analysis so that the associated emotions are perceived. The features extracted by this system are occupied area, body displacement and overlap/interactions between people.

This system is divided into several parts. Initially, the points of interest are extracted from the body of each person present in the image under analysis. Then, using these same points of interest, the characteristics are extracted, starting with the occupied area by the body of the person in question, displacement of body parts over time and, finally, overlaps between the people present in the images over time.

A. Points of Interest

Points of interest reflect each joint of the human body being detected up to 18 joints by OpenPose library. In this way, the initial step of the developed system starts by processing the video files, frame by frame, through this library. As an output of the processing, results a set of JavaScript Object Notation (JSON) files relative to the points of interest detected in each frame, as well as a video where the detected postures are presented in each frame of the same. JSON files are structured through value lists with pairs (key/value), making it simple, fast and efficient to process the information contained in them.

Each file contains all points of interest relative to the persons found in the respective frame, in form of x, y, c, x, y, c, ..., where x and y represent the coordinates of each joint and c represents the percentage of confidence that the algorithm has on the previous coordinates, in the order defined in the library and which is shown in Figure 10.

An excerpt of one of the files generated by the library is shown in Figure 11, where the result can be verified for the first person detected in the respective frame this file.

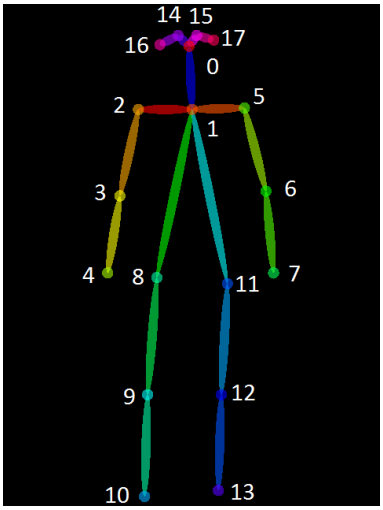

Figure 10. Image representative of the order of points of interest defined by the OpenPose library [39].

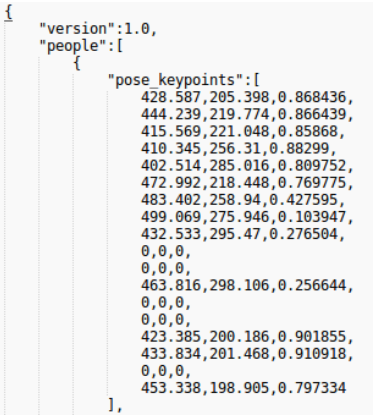

Figure 11. Excerpt from a JSON point of interest file generated by the OpenPose library.

B. Occupied Area

Using the JSON files obtained above, containing the values of the 2D positions of the main parts of the human body, we obtain the maximum and minimum x and y values of each person and, by using these values, it is possible to calculate the width, height and, consequently, the occupied area by the person. This procedure is done for all people in the frame under review. After obtaining these maximum and minimum values, bounding boxes around each person are drawn, so that it is possible to verify the results obtained by visual inspection.

Graphs are created to visually observe the evolution of occupied areas over time by each person. These graphs will be shown and explained in the Results section.

C. Displacements

The displacement values are calculated for each part of the body, individually. The body parts (vectors) are defined by a pair of points of interest representing two connected body joints, in a total of 13 vectors, as shown in Figure 12.

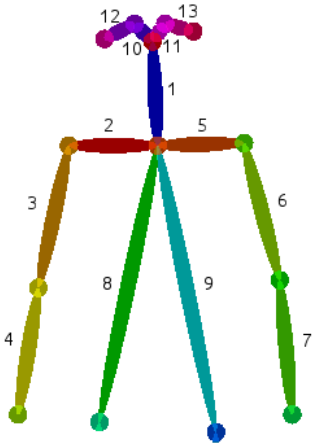

Figure 12. Representation of the vectors associated with the 13 parts of the study body.

The calculation is done frame by frame, analyzing the current and previous frame. In this way, it is possible, by calculating the Euclidean distance, to determine the displacement, relative to time, of each vector. However, there is the problem of failure to detect some points of interest in some of the frames. If this problem occurs, then the value of the last detection of the point of interest in question is saved and the calculation made once it is detected again. Thus, it is still possible to calculate the displacement velocity by dividing the distance by the value of the interval of frames in which the vector was not found.

The values relative to the distance are stored in files respecting the increasing order of the vector number, finally writing the frame number and the area occupied by the person. If it is not possible to calculate any of the vectors in the frame in question, then a '-' is placed in its position. With these values graphs that allow to visually observe the evolution of the displacements over time are generated. These graphs will be shown and explained in Section VI.

D. Overlays

Once persons are identified, the overlaps between them are studied, in order to gauge a possible interaction between the elements of the group. Detection of overlaps involves the estimation of a depth value, since all values obtained are 2D. Thus, a cycle is made that allows to perform a check between all possible combinations of people, two by two. The estimate is based on the minimum and maximum values of x and y of the two persons to be compared, as well as the values of x and y of the upper limbs of each.

This estimation starts by checking for overlap between the two bounding boxes. If it does not exist then it is assumed, immediately, that there is no overlap. If it exists, a possible overlap is assumed. However, there are no guarantees yet, since a precise depth estimate has not yet been made. To increase the certainty in the estimation it is verified that the distance between the maximum values of y of the two bounding boxes is relatively small. If not, then it is assumed that the two persons are not in the same plane at z and, consequently, assume a non-overlap / interaction. If so, then the level of confidence in relation to the overlap increases, and there is already an estimate of depth. To further increase certainty in the detection the opening of the upper limbs of each person is verified and, if these openings are identical, it is verified if the area of the two bounding boxes is identical. If it is not, the estimate is ended with a possible low confidence interaction. If it is, the confidence increases, and it is checked again if the distance in y of the two bounding boxes is even closer, in order to increase even more the level of confidence. In this way it is possible to make a relatively precise depth estimate so that possible overlaps/interactions are detected. Figure 13 graphically represents the detection of an overlap.

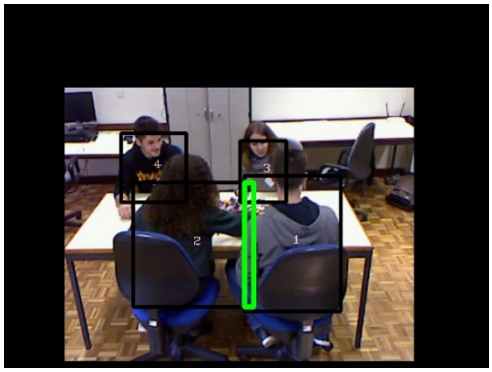

Figure 13. Graphic representation of an overlay.

With the overlaps found, heatmaps are created that allow to perceive visually the regions where there was a greater number of interactions / overlays. A heatmap is a graphical representation of data, where values are represented by a color scale.

The creation of the heatmap is done through a matrix that is zeroed and has the same size as the images under study. Whenever there is a possible overlap, all values relating to the overlapping region in the matrix are incremented. At the end all values are converted to a scale from 0 to 255 where a

colorMap is used to distinguish zones with more interactions (red) from zones with less interactions (blue). In the Results section, examples of heatmaps obtained will be presented.

VI. RESULTS

To test the developed system, videos obtained through a study at the Department of Education and Psychology of the University of Aveiro, entitled "Conflict, emotions and efficacy in work teams: an experimental study" were used. This study consists of several sessions, where each session involves 4 participants in interaction around a table.

The experimental phase of each session consists of two steps, which involve two construction tasks (with legos) that must be performed in a group. The first task focuses on the construction of a horse, taking the participants 15 minutes to be performed. The second task is to construct a spacecraft, and the participants should begin by presenting and discussing the individual ideas for the next plan together, in order to decide how to proceed, with 5 minutes to do so. After the 5 minutes of discussion, 20 minutes are given for the construction of the spacecraft.

The experiments carried out can be related to two types of conditions: the control condition, where no conflict is induced and the goal is to evaluate how the emergence of conflicts and the associated emotions influence the functioning of the work teams; the condition of task conflict, where for the second task are given dubious and complex instructions.

In the results presented below, for the characteristics: occupied area and displacements were generated graphs that are divided in two parts, where the first part refers to the first phase of the experiment and the second part referring to the second phase of the experiment. Figure 14 represents the division made. In this way it is possible to analyze the behavior change of each participant according to the condition tested.

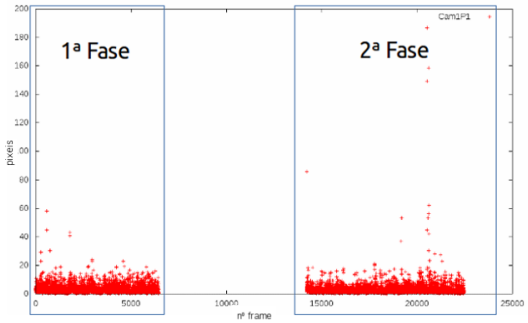

Figure 14. Division of the graphs obtained by the two phases of the experiment.

A. Occupied Area

To observe the variation of occupied area by each participant, graphs were created with the data obtained through the processing of the videos obtained. Thus, it is possible to verify if participant changed their posture between the two phases of the experiment, occupying more or less space. Figures 15, 16, 17 and 18 show the graphs relating to the area of the participants during an experiment in which it was not induced any conflict.

Figures 19, 20, 21 and 22 show the graphs relating to the area of the participants during an experiment in which the task conflict was induced. By observing these graphs, it is verified that the variation of the area between the two phases of the experiment is, on average, greater when there is an induction of conflict.

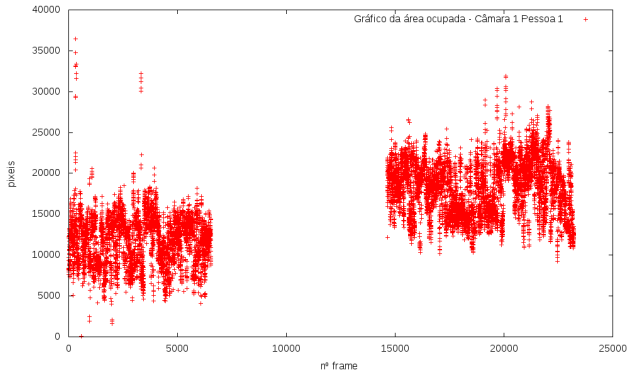

Figure 15. Graph relative to the occupied area by the person with id 1 in an experiment without inducing conflict (12/13/2017).

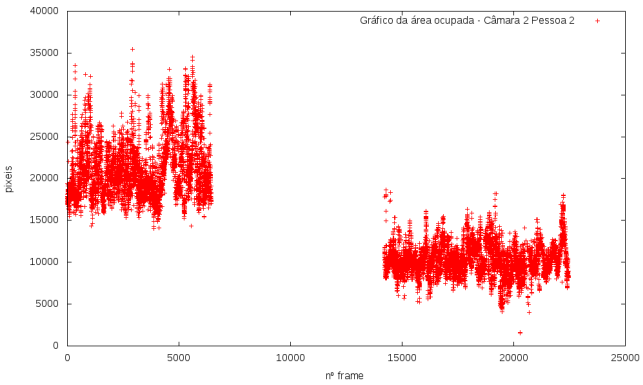

Figure 16. Graph relative to the occupied area by the person with id 2 in an experiment without inducing conflict (12/13/2017).

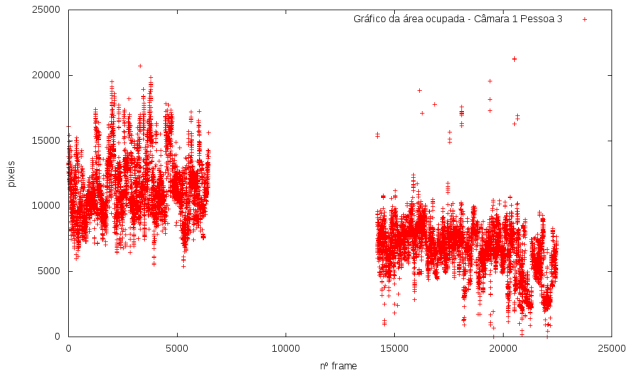

Figure 17. Graph relative to the occupied area by the person with id 3 in an experiment without inducing conflict (12/13/2017).

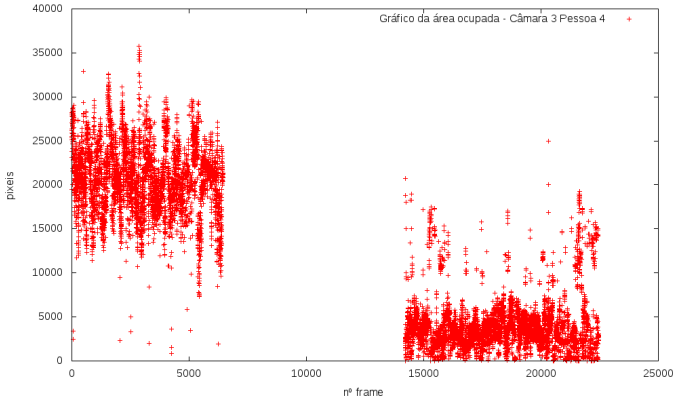

Figure 18. Graph relative to the occupied area by the person with id 4 in an experiment without inducing conflict (12/13/2017).

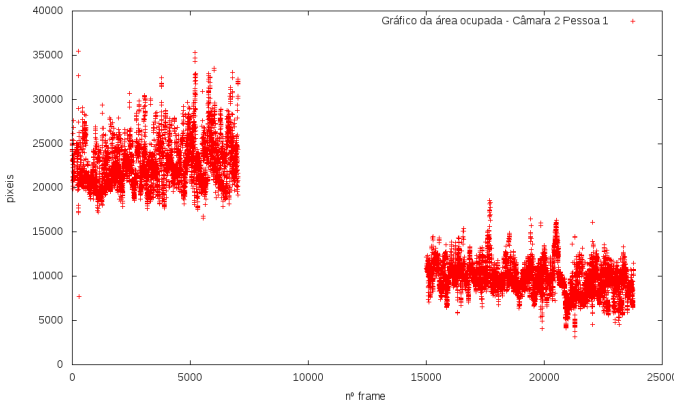

Figure 19. Graph relative to the occupied area by the person with id 1 in an experiment with induction of task conflict (12/15/2017).

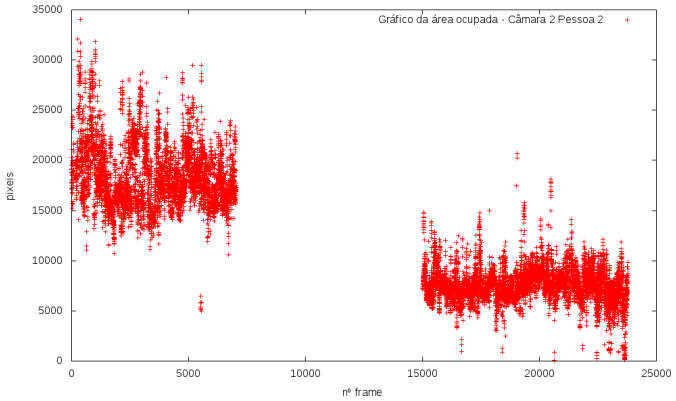

Figure 20. Graph relative to the occupied area by the person with id 2 in an experiment with induction of task conflict (12/15/2017).

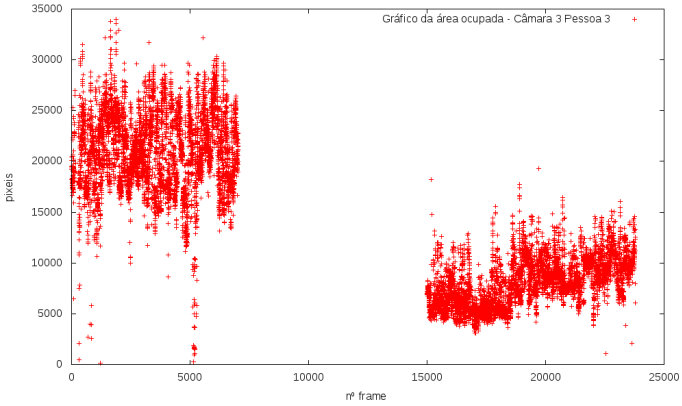

Figure 21. Graph relative to the occupied area by the person with id 3 in an experiment with induction of task conflict (12/15/2017).

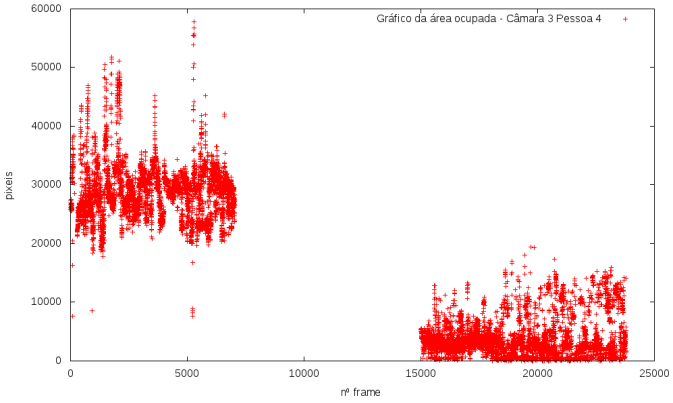

Figure 22. Graph relative to the occupied area by the person with id 4 in an experiment with induction of task conflict (12/15/2017).

B. Displacements

To verify the evolution of the displacements, we draw graphs that demonstrate the values of the displacements in each frame. These values can be related to each vector or the average of all vectors found.

Figures 23, 24, 25 and 26 show the graphs relating to the movement of participants in an experience where was not induced any conflict. Figures 27, 28, 29 and 30 show the graphs relating to the movement of the participants during an experiment in which the task conflict was induced. By observing these graphs, it is observed that, although small, there is a greater difference of movement between the two phases of the experience when there is induction of conflict.

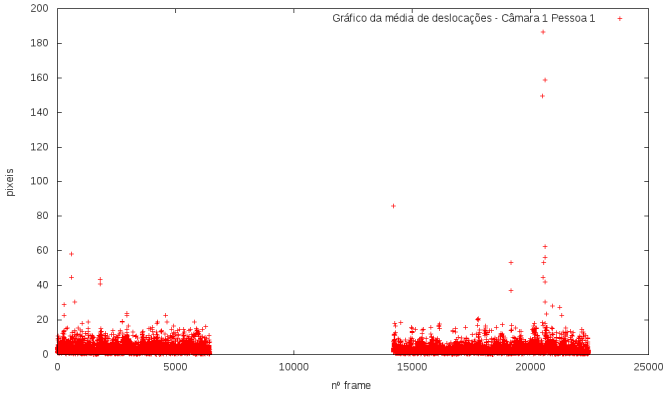

Figure 23. Graph on the average displacements of the person with id 1 in an experiment without inducing conflict (12/13/2017).

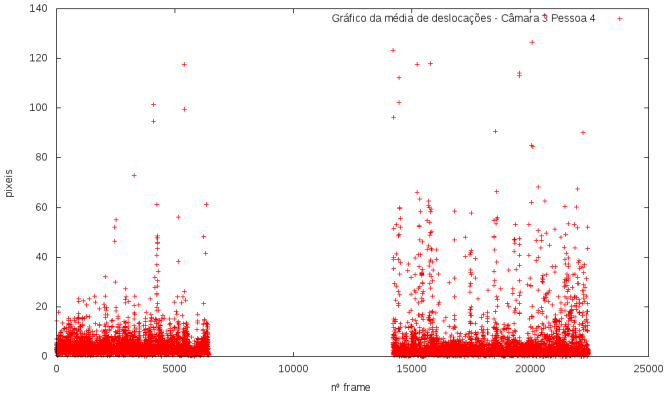

Figure 26. Graph on the average displacements of the person with id 4 in an experiment without inducing conflict (12/13/2017).

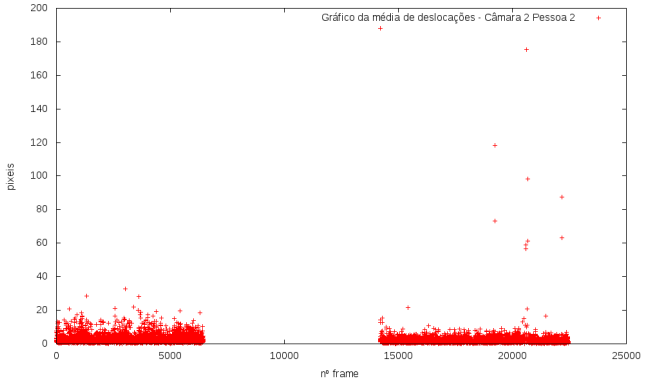

Figure 24. Graph on the average displacements of the person with id 2 in an experiment without inducing conflict (12/13/2017).

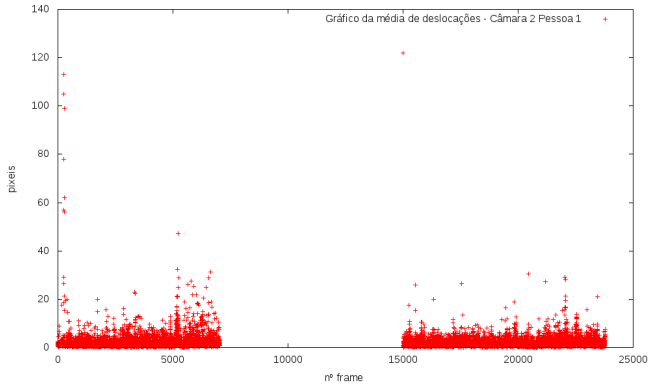

Figure 27. Graph on the average displacements of the person with id 1 in an experiment with induction of task conflict (12/15/2017).

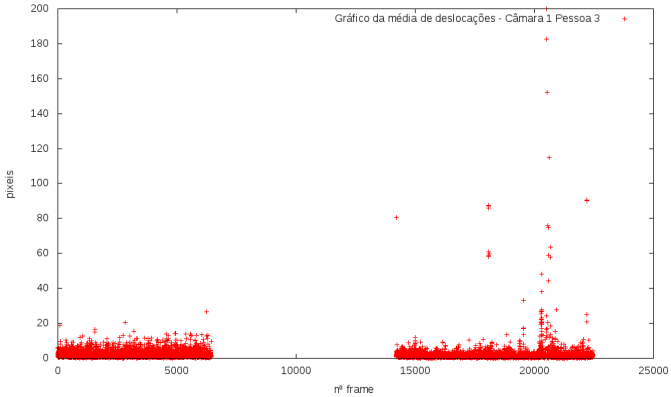

Figure 25. Graph on the average displacements of the person with id 3 in an experiment without inducing conflict (12/13/2017).

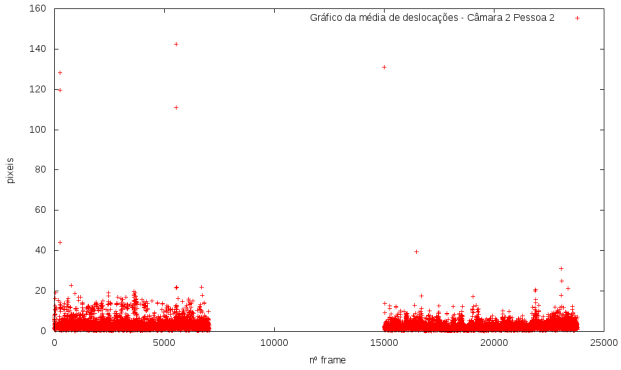

Figure 28. Graph on the average displacements of the person with id 2 in an experiment with induction of task conflict (12/15/2017).

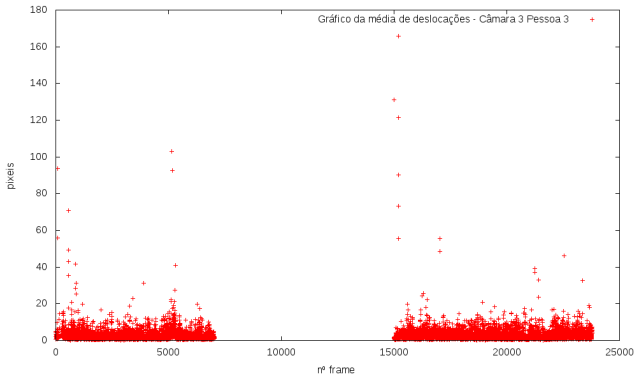

Figure 29. Graph on the average displacements of the person with id 3 in an experiment with induction of task conflict (12/15/2017).

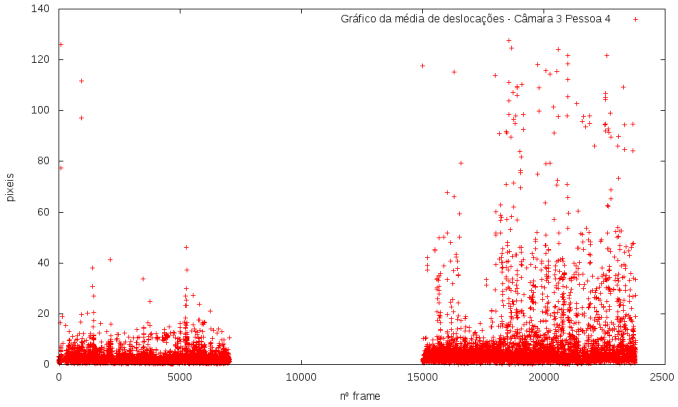

Figure 30. Graph on the average displacements of the person with id 4 in an experiment with induction of task conflict (12/15/2017).

C. *Overlays*

Finally, to be able to observe the level of interaction between participants, heatmaps were created in each experiment. Thus, it is possible to verify which participants have more interaction, i.e., it can be shown which were the ones that played a more important role during the same. By observing each heatmap, it is possible to quickly verify the areas with the highest interaction, although some minimal errors occur in the identification of people during the detection process.

For example, when looking at Figures 31, 32, and 33, it turns out that people 1 and 2 were the ones that most interacted. Already people 3 and 4 had a much lower number of interactions. Comparing Figures 31, 32, and 33, which relate to an experimental condition where no conflict was induced, with Figures 34, 35, and 36, which relate to a condition where task conflict was induced, it can be observed that the number of interactions/overlaps is higher when there is induction of task conflict.

Figure 31. Heatmap relative to camera 1 in an experiment without inducing conflict (12/13/2017).

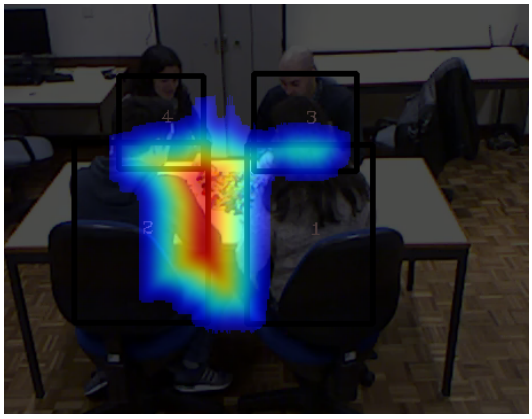

Figure 32. Heatmap relative to camera 2 in an experiment without inducing conflict (12/13/2017).

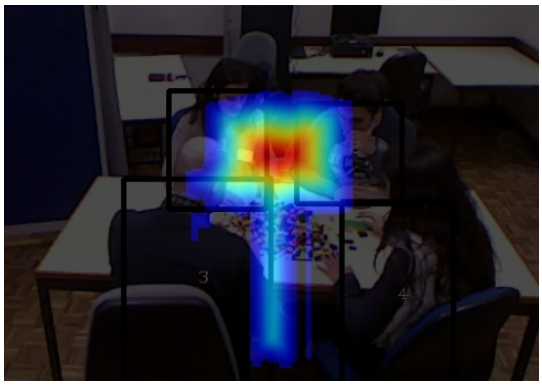

Figure 33. Heatmap relative to camera 3 in an experiment without inducing conflict (12/13/2017).

Figure 34. Heatmap relative to camera 1 in an experiment with task conflict induction (12/15/2017).

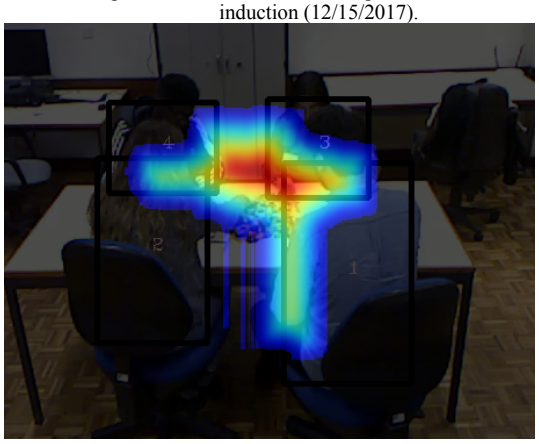

Figure 35. Heatmap relative to camera 2 in an experiment with task conflict induction (12/15/2017).

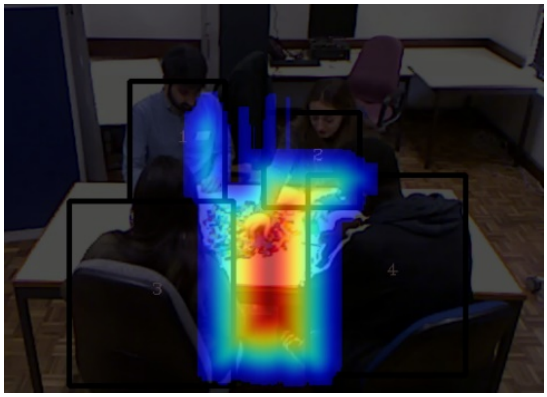

Figure 36. Heatmap relative to camera 3 in an experiment with task conflict induction (12/15/2017).

VII. Conclusion

In the presented state of the art, it is possible to recognize the lack of systems for the detection and classification of emotion systems from human body postures, as well as the difficulties associated to the already existing systems. However, there are many image sensor's alternatives, allowing to guide the system to several types of solutions, from skeleton detection based on distance image to the detection based on RGB image.

The Human Posture Detection algorithms research work [32] presents the algorithm with better results at all levels, which is possibly what will be used in the development of a system to allow the recognition of emotions from human body postures. This solution is not only optimal for its simplicity in terms of image, but also for its good results in detecting postures in groups of people. However, for this algorithm to work properly, it is necessary to have specific and expensive hardware, due to the parallel computing used and the GPU calculation performed.

With the developed system it was possible to perceive that characteristics can be extracted allowing to assess the postures and associated emotions. As an example, looking at a graph relative to the area occupied by a person during an experiment can be concluded if the same assumed a more contracted or expansive posture, which is associated with negative and positive emotions, respectively. Looking further at a graph relating to movement of the joints of the body it is possible to deduce if the person is more or less confident. Importantly, thought, future studies should further develop emotion categorization via body postures regarding other emotional conditions (e.g., in fear contexts), using technological solutions. The use of such ubiquitous solutions for a variety of emotions may be of great value to determine the individual emotional experience more objectively and, ultimately, enable adaptations in the context (e.g., a classroom) to enhance well-being, particularly when these reflect a negative valence.

Finally, through the observation of the created heatmaps, the level of interactions between the participants could be directly related to the amount of movement and the posture assumed by each one, that is, if individual A moved very little, so it will have few interactions. On the other hand, if two individuals B and C, which are side by side, have moved a lot, then the interactions in that zone will be high.

VIII. Acknowledgment

This work was partially funded by FEDER (Programa Operacional Factores de Competitividade - COMPETE), by National Funds through the FCT - Foundation for Science and Technology in the context of the project UID/CEC/00127/2013 and by the Integrated Programme of SR\&TD "SOCA" (Ref. CENTRO-01-0145-FEDER-000010), co-funded by Centro 2020 program, Portugal 2020, European Union, through the European Regional Development Fund.

IX. References

[1] B. Barbosa, A. J. R. Neves, S. C. Soares, I. D. Dimas, C. Ua, and D. De Educação, "Analysis of Emotions from Body Postures Based on Digital Imaging," in. Proc. *SIGNAL 2018*, The Third International Conference on Advances in Signal, Image and Video Processing, pp. 73–78, 2018.

[2] B. de Gelder, A. W. de Borst, and R. Watson, "The perception of emotion in body expressions," *Wiley Interdiscip. Rev. Cogn. Sci.*, vol. 6, no. 2, pp. 149–158, 2014.

[3] A. P. Atkinsonô, W. H. Dittrichô, A. J. Gemmell, and A. W. Young, "Emotion perception from dynamic and static body expressions in point-light and full-light displays," *Perception*, vol. 33, pp. 717–746, 2004.

[4] W. H. Dittrich, T. Troscianko, S. E. G. Lea, and D. Morgan, "Perception of Emotion from Dynamic Point-Light Displays Represented in Dance," *Perception*, vol. 25, no. 6, pp. 727–738, Jun. 1996.

[5] E. Babad, "Teaching and nonverbal behaviour in the classroom," in *International Handbook of Research on Teachers and Teaching*, Boston, MA: Springer US, 2009, pp. 817–827.

[6] H. A. Elfenbein, J. T. Polzer, and N. Ambady, "Team Emotion Recognition Accuracy and Team Performance," *Research on Emotion in Organizations*, vol. 3. pp. 87–119, 2007.

[7] F. Loi, J. G. Vaidya, and S. Paradiso, "Recognition of emotion from body language among patients with unipolar depression," *Psychiatry Res.*, vol. 209, no. 1, pp. 40–49, Aug. 2013.

[8] T.-L. L. Le, M.-Q. Q. Nguyen, and T.-T.-M. T. M. Nguyen, "Human posture recognition using human skeleton provided by Kinect," *2013 Int. Conf. Comput. Manag. Telecommun.*, pp. 340–345, 2013.

[9] Z. Zhang, Y. Liu, A. Li, and M. Wang, "A Novel Method for User-Defined Human Posture Recognition Using Kinect," *Int. Congr. Image Signal Process.*, pp. 736–740, 2014.

[10] C. W. Chang, M. Da Nian, Y. F. Chen, C. H. Chi, and C. W. Tao, "Design of a Kinect Sensor Based Posture Recognition System," *2014 Tenth Int. Conf. Intell. Inf. Hiding Multimed. Signal Process.*, pp. 856–859, 2014.

[11] M. Raca and P. Dillenbourg, "Classroom Social Signal Analysis," *J. Learn. Anal.*, vol. 1, no. 3, pp. 176–178, 2014.

[12] C. Thurau and V. Hlavac, "Pose primitive based human action recognition in videos or still images," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[13] Tao Zhao and R. Nevatia, "Tracking multiple humans in complex situations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1208–1221, Sep. 2004.

[14] R. Gonzalez and R. Woods, "Digital image processing and computer vision," *Comput. Vision, Graph. Image Process.*, vol. 49, no. 1, p. 122, Jan. 1990.

[15] M. Lyra, A. Ploussi, and A. Georgantzoglou, "MATLAB as a Tool in Nuclear Medicine Image Processing," *MATLAB - A Ubiquitous Tool Pract. Eng.*, no. October 2011, pp. 477–500, 2011.

[16] G. Borenstein, *Making Things See: 3D Vision with Kinect, Processing, Arduino, and MakerBot*. 2012.

[17] "Free How to Photoshop Tutorials, Videos & Lessons to learn Photoshop training | Photoshop Course." [Online]. Available: http://www.we-r-here.com/ps/tutorials/. [Accessed: 28-Nov-2017].

[18] "Naushadsblog." [Online]. Available: https://naushadsblog.wordpress.com/. [Accessed: 28-Nov-2017].

[19] N. Blanc, "CCD versus CMOS - has CCD imaging come to an end?," *Photogramm. Week 2001*, pp. 131–137, 2001.

[20] "Wikimedia Commons." [Online]. Available: https://commons.wikimedia.org/wiki/Main_Page. [Accessed: 28-Nov-2017].

[21] "Photography tips and tricks, Equipment, Photography News, Photography Books, Tutorial, and Lighting - OneSlidePhotography.com." [Online]. Available: http://oneslidephotography.com/. [Accessed: 28-Nov-2017].

[22] W. K. Wong, P. N. Tan, C. K. Loo, and W. S. Lim, "An effective surveillance system using thermal camera," *2009 Int. Conf. Signal Acquis. Process. ICSAP 2009*, pp. 13–17, 2009.

[23] "FLIR Systems | Thermal Imaging, Night Vision and Infrared Camera Systems." [Online]. Available: http://www.flir.eu/home/. [Accessed: 28-Nov-2017].

[24] T. Sosnowski, G. Bieszczad, and H. Madura, "Image Processing in Thermal Cameras," Springer, Cham, 2018, pp. 35–57.

[25] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07–12–June, pp. 1037–1045, 2015.

[26] L.-J. Ferrato and K. W. Forsythe, "Comparing Hyperspectral and Multispectral Imagery for Land Classification of the Lower Don River, Toronto," *J. Geogr. Geol.*, vol. 5, no. 1, pp. 92–107, 2013.

[27] "What is the difference between multispectral and hyperspectral imagery? - eXtension." [Online]. Available: http://articles.extension.org/pages/40073/what-is-the-difference-between-multispectral-and-hyperspectral-imagery. [Accessed: 28-Nov-2017].

[28] M. Aboras, H. Amasha, and I. Ibraheem, "Early detection of melanoma using multispectral imaging and artificial intelligence techniques Early detection of melanoma using multispectral imaging and artificial intelligence techniques," *Http://Www.Sciencepublishinggroup.Com*, vol. 3, no. November 2016, p. 29, 2015.

[29] S. B. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor - System description, issues and solutions," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2004–Janua, no. January, 2004.

[30] P. Zanuttigh, C. D. Mutto, L. Minto, G. Marin, F. Dominio, and G. M. Cortelazzo, *Time-of-flight and*

structured light depth cameras: Technology and applications. 2016.

[31] G. Calin and V. O. Roda, "Real-time disparity map extraction in a dual head stereo vision system," *Lat. Am. Appl. Res.*, vol. 37, no. 1, pp. 21–24, 2007.

[32] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1302–1310.

[33] Jungong Han, Ling Shao, Dong Xu, and J. Shotton, "Enhanced Computer Vision With Microsoft Kinect Sensor: A Review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, Oct. 2013.

[34] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE Multimed.*, vol. 19, no. 2, pp. 4–10, 2012.

[35] "Kinect for Windows Sensor Components and Specifications." [Online]. Available: https://msdn.microsoft.com/en-us/library/jj131033.aspx. [Accessed: 02-Dec-2017].

[36] "OpenPose - Realtime Multiperson 2D Keypoint Detection from Video | Flintbox." [Online]. Available: https://cmu.flintbox.com/public/project/47343/. [Accessed: 03-Dec-2017].

[37] E. Insafutdinov *et al.*, "ArtTrack: Articulated Multi-person Tracking in the Wild," Dec. 2016.

[38] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model," May 2016.

[39] "GitHub - CMU-Perceptual-Computing-Lab/openpose: OpenPose: Real-time multi-person keypoint detection library for body, face, and hands estimation." [Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/openpose. [Accessed: 17-May-2018].