# Human Body Posture Detection in Context: The Case of Teaching and Learning Environments

Rui Sacchetti, Tiago Teixeira,
Bruno Barbosa, António J. R. Neves
DETI / IEETA
University of Aveiro
3810-193 Aveiro, Portugal
Email: {ruisacchetti, tiagomaioteixeira, 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*—This paper describes an approach to detect and classify human posture in an individual context, more precisely in a classroom ambience. The posture can be divided into two main groups: "Confident/Not Confident", aiming for the teacher's posture evaluation, and "Interested/Not Interested", targeted for the students. We present some relevant concepts about these postures and how can they be effectively detected using the OpenPose library. The library returns the main key points of a human posture. Next, with TensorFlow, an open-source software library for machine learning, a deep learning algorithm has been developed and trained to classify a given posture. Lastly, the neural network is put to the test, classifying the human posture from a video input, labeling each frame. The experimental results presented in this paper confirm the effectiveness of the proposed approach.**

***Keywords - Body language; Human Postures; Computer Vision; Digital Camera; Machine Learning.***

## I. Introduction

The advancement of computers and new technologies plays a key role in creating systems capable of better interacting with humans, which leads to an increasing number of systems that can analyze, classify and predict human behavior. Emotions are at the core of most of our overt behavior and plays a key role in socio-communicative interactions [1]. Facial expressions are exceptionally powerful nonverbal means to convey information about emotions [2] and have been the focus of an enormous amount of research for several decades.

More recently, studies have also been devoting a great deal of attention to human body postures, as they also express emotions, while adding important cues on behavioral intentions, which is critical for functional social interactions [3]. For instance, one may infer from Figure 1 that the person on the left seems to be more insecure/less confident than the person on the right.

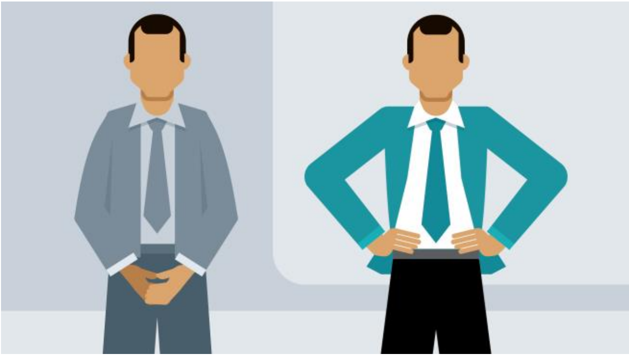

Figure 1. Body language differences: insecure to the left and confident to the right [4].

The possibility to assess emotional and motivational dimensions in educational contexts, by means of machine learning techniques, was the main motivation of the current work. More specifically, within a classroom context, our goal was to test the feasibility of categorizing the speaker, either as confident or non-confident (see Figure 1), and the audience, with either interested or non-interested postures (as depicted in Figure 2). An effective two-way interaction between the audience and the speaker is deemed as crucial to engage the audience, which is why implementing non-invasive technological tools may provide meaningful improvements in educational settings.

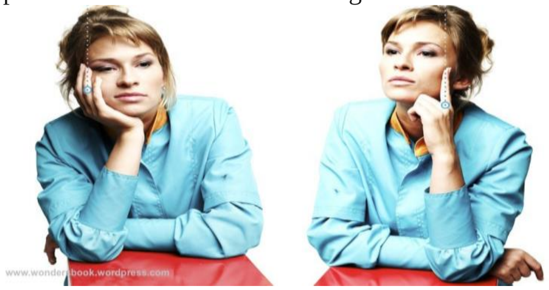

Figure 2. Body language differences: Not Interested to the left and Interested to the right [5].

ISBN: 978-1-61208-638-5

For the development of the proposed classifier, we used a state-of-the-art algorithm to calculate human postures: the OpenPose library [6][7]. Using a common RGB camera, it is possible to obtain the main key points of an individual on the scene. Then, after obtaining this information, we use a deep learning image classifier to train the neural network based on the TensorFlow library [8][9]. Later, we use the trained model to classify each frame of a given video, frame by frame, and label the posture of the individual. These procedures will be described in more detail in the next section. Finally, Section III presents the experimental results obtained, followed by a conclusion in Section IV.

## II. PROPOSED APPROACH

This work was developed in well-defined gradual stages. In this section, we describe all the phases and technologies in detail. We begin with a brief description of the OpenPose library [6][7], how it was used to extract the key points, and the results obtained. Next, TensorFlow [8][9] machine learning library is described, as well as another library, Keras [10][11], which runs on top of TensorFlow and speeds up the development of deep learning models. Finally, we discuss how the generated models of deep learning were used to classify postures.

### A. *OpenPose*

OpenPose is an open-source C++ library for detecting key points in human postures [6][7]. It was recently launched (2017) and, due to its potential, has been widely used for different purposes. In this work, in order to provide clear images about the human posture to the deep learning algorithm - that is, without noise or other elements that might compromise the efficiency of the algorithm - this library was used in order to obtain the best learning degree possible.

The library can detect 15 or 18 body key points, 21 key points per hand and 70 face key points. It can detect multiple individuals in one scene. However, with more than one person in the scene, the speed of detection is greatly reduced. It uses deep learning algorithms for better detection of the person's key points, using the Caffe framework [12].

It can be used with command-line demo, C ++ wrapper or C ++ API, and can receive as input images, videos, webcam images or IP cameras. The output of this library can be varied due to the number of flags that can be used, for example: include hands / face detection, just represent the key points (no background), save results to images (video frames), save various key points identified in files, etc. Consequently, this library and its features require a computer with large computational and parallel processing capabilities, as well as the installation of specific software. The computer used for the development and test of the proposed approach has the following characteristics:

Software:

- OpenCV (version 3.3.1);
- Caffe (version *custom*);
- CUDA (version 8.0);
- CuDNN (version 5.1).

Hardware:

- Nvidia 1080 TI (11GB *frame buffer*);
- 32Gb RAM;
- Intel i7 with 8 *cores*.

### B. *OpenPose Results*

The first step was to build the database of "Confident" postures. For this, several videos were recorded in which the subject showed himself with a positive and strong attitude (see Figure 3). Quickly, and with the naked eye, it was possible to detect some characteristics about the key points that a posture of the type "Confident" returns, for example:

- The level of the shoulders forms a line perpendicular to the line of the spine.
- The subject's head is always with a degree of 90°, or greater, in relation to the spine.
- The arms are slightly further away from the body.

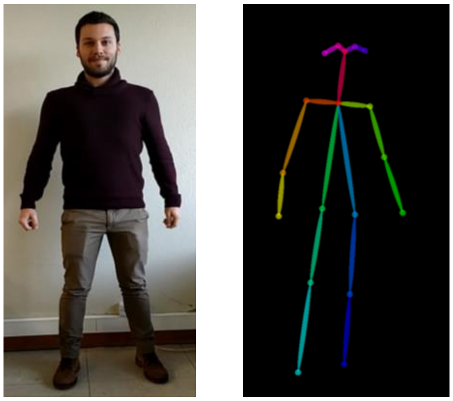

Figure 3. Schematic representation of key points provided by the OpenPose library.

Three individuals participated in the video samples, where each one recorded at least one video enhancing one of the 4 labels. In total, 1002 pictures tagged "Confident" were created. The same steps were repeated for a "Not Confident" posture where 1220 images were generated.

The next step was to construct the "Interested / Not Interested" type database. Several videos were recorded in which the subject enhanced positions that showed interest - like a more advanced posture, arms on the table or the head in the alignment of the body.

In this type of posture, the key points of the face and hands are quite expressive and quite strong indicators of the subject's posture. Hence, with the use of --*hand* and --*head* functionalities, the OpenPose returns the key points of the face and hands. The main goal was to obtain the best accuracy possible in the deep learning algorithm (see Figure 4).

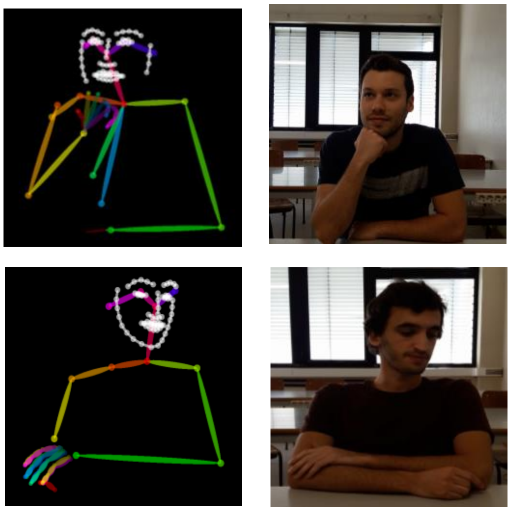

Figure 4. Representation of the key points: “Interested” above, “Not Interested” below.

In many generated frames, the position of the hands was not correctly determined. Therefore, to obtain the best possible precision in TensorFlow, images with only the --head functionality were also extracted. In addition, it was decided to do another extraction but only with the basic (18) key points. This way, we studied 3 different cases for the type “Interested / Not Interested" posture: basic points, face and hands; basic points and face and only basic points. These results were stored in different databases. The objective was to study which case obtained the better precision values in the classification model. 1270 images labeled "Interested” and 1335 "Not Interested" images were obtained, for each type, one of the three cases that are about to be analyzed.

## C. *TensorFlow*

TensorFlow is an open-source library for machine learning, more specifically for deep learning, developed by Google in 2015 [8][9]. This library was chosen because it allows the development of classifiers in an easier way than the other options considered, not compromising the quality of the results. The Keras library [10][11] was also used since it eases the workflow of creating a neural network.

Because our training set are digital images, we decided to use the neural network class, i.e., Convolutional Neural Network (CNN). The great advantage of using CNN is that it does not require pre-processing when compared to other image classification algorithms. Thus, it was possible to train a network without prior knowledge.

The TensorFlow library is available in Python and C++, whereas Keras is only available in Python. We developed our solution in Python programming language to do image classification.

The first step was to develop the classifier training software, obtaining the model that would classify the images. The biggest challenge was to find the most accurate parameters that would lead to good classifier accuracy. Better results were experimentally obtained when using 25 epochs and a batch size of 100 images.

When training the "Confident / Not Confident" model, the precision results obtained were about 93% for validation and 87% for the test phase, which were quite satisfactory. Figure 5 shows the performance of the classifier when trained.

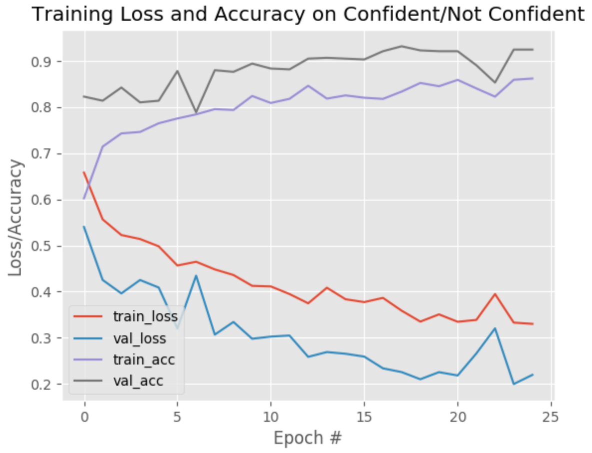

Figure 5. Results obtained when training the "Confident / Not Confident" classifier.

Then, the models of "Interested / Not Interested" were trained, for the three different cases, in search of the best results.

For the case with face and hands, the results are presented in Figure 6.

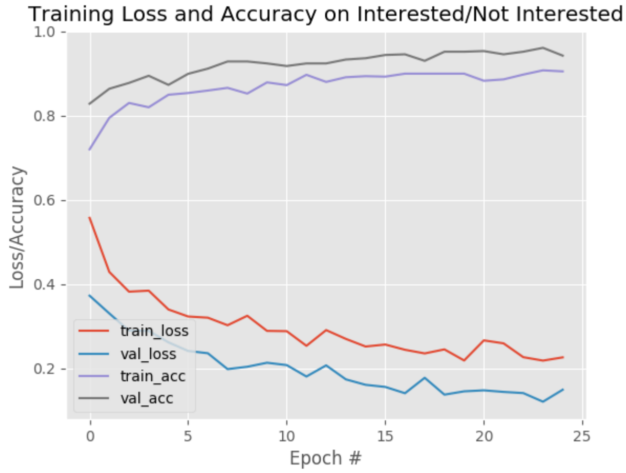

Figure 6. Results obtained when training the "Interested / Not Interested" classifier, for the head and hands case.

As shown in Figure 6, the precision values were steadily increasing, reaching a maximum validation value of 94%.

For the case that only considers the hands, the results obtained are presented in Figure 7.

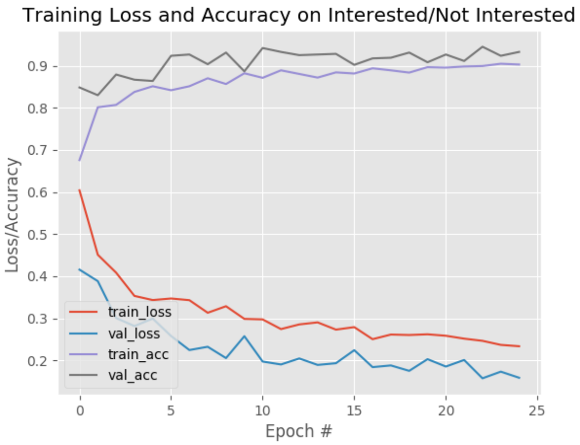

Figure 7. Results obtained when training the "Interested / Not Interested" classifier, for the hands only case.

Finally, for the database of images with only the basic points, the obtained results are presented in Figure 8.

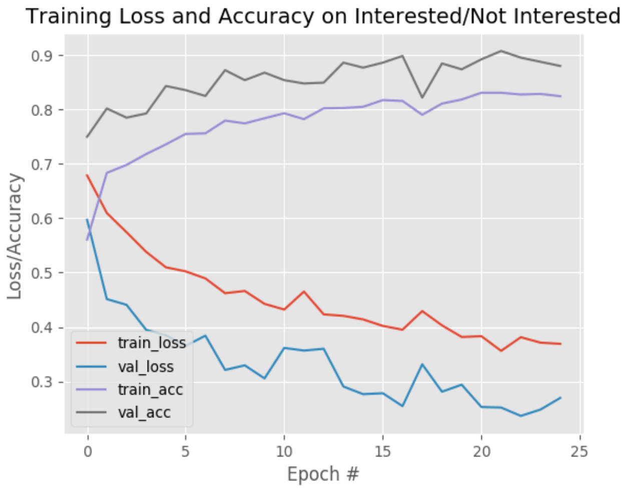

Figure 8. Results obtained when training the "Interested / Not Interested" classifier, for basic key points case.

It is possible to conclude that the model with face and hands obtained a better precision when compared with the others, obtaining a final precision of, approximately, 94%. Thus, this will be the model used for the classification phase.

The next step was to develop the classifier that allows us to obtain a label - “Confident / Not Confident” or "Interested/ Not Interested", with the input of a video and a previously trained model. This program is described in the following subsection.

### D. Classifier

The developed classifier loads the classification model and opens the provided video. Then, frame by frame, the program questions the model and gets the posture label for that frame. The program returns the subject's posture classification, not only based on the current frame, but also with the previous frames.

Since the posture of an individual does not change frequently in a short period of time, and to eliminate false positives, we decided to create a filter that rules out false positives. First, we implemented a sliding window that follows the behavior of the subject in the last seconds. For the attribution of the final label, we considered the average of all postures analyzed in this time window. Then, to eliminate cases in which the model does not have a high degree of certainty about the analyzed frame, a threshold was created that only considered frames that contained a reliable degree of certainty. After some tests, we decided that the certainty percentage of a given label should be greater than 65% to be considered.

## III. Experimental Results

To verify the effectiveness of the developed classifier for the "Confident / Not Confident" postures, a new video was recorded in which the subject initially presented himself with a more contracted posture. This video was not used during the training phase for the deep learning algorithm. Any assigned ratings were based on already existing databases.

Throughout the video, the subject changed his posture to a more expansive posture, reflecting a more confident attitude. Finally, it changed again to a more contracted posture. It was expected that the label assigned to each change was the correct one.

Observing the obtained output (see Figure 9, 10), we conclude that the developed classifier properly labeled the subject's posture. In the first seconds (2-3s), while the sliding window is not filled - and there are no certainties of about the posture - no label is assigned.

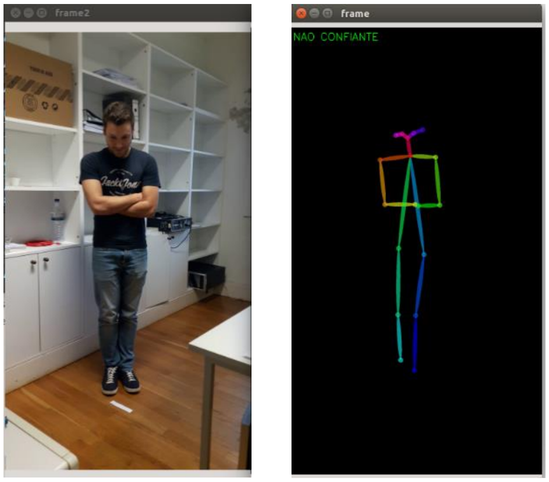

Figure 9. Results obtained for a "Not Confident" posture. In image to the right, the label assigned is at the upper left corner.

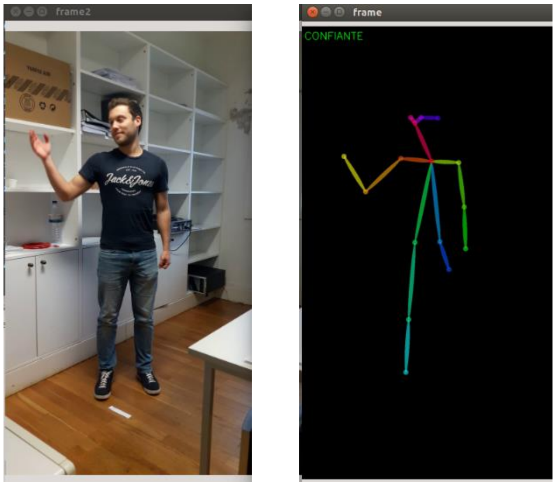

Figure 10. Results obtained for a "Confident" posture.

Due to decisions taken to prevent false positives, the classifier became quite robust. It took some testing to adjust the correct size of the sliding window so that the program did not become too strict and did not require a lot of time to obtain the posture confirmation. At the moment of this writing, the sliding window had a size of 25 frames.

To test the detection of sitting postures, i.e., "Interested / Not Interested", the same procedure as the previous classifier were followed. A video was recorded in which the subject enhanced a less interested posture (looks at the sides, up, has the head rested in one hand) (see Figure 11). As the video continued, the subject's posture changed to an "Interested" posture (see Figure 12) and after some time this posture was changed back to "Not Interested". In this test case, the classifier also labeled correctly the posture presented throughout the video.

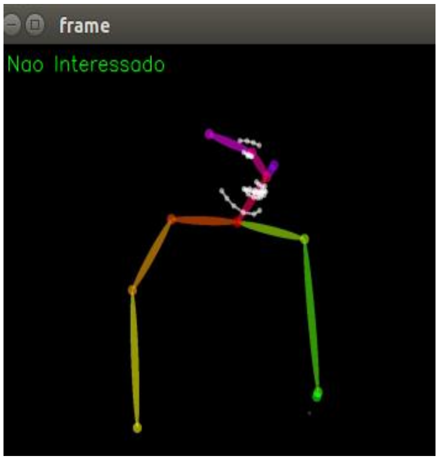

Figure 11. Results obtained for a "Not Interested" posture.

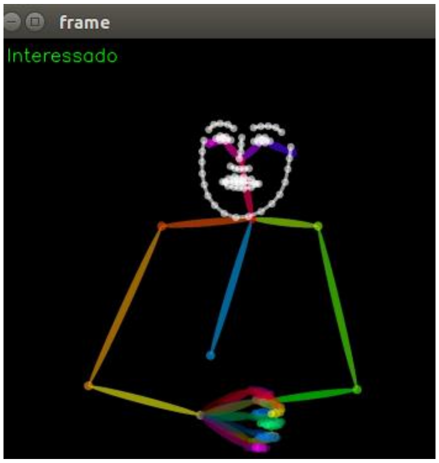

Figure 12. Results obtained for an "Interested" posture.

## IV. Conclusion and Future Work

The comprehension of body language is an area that is in great development. However, despite the positive results obtained in this work, this is a very complex research domain, mostly because human emotions and their body language can't be classified in a binary way. It is not certain that a person who is speaking to an audience with a "bad" posture is less confident or less relaxed. There are other indicators that are important to this classification. This work can be used as a small increment to this theme. In future work, this type of posture classification can be used simultaneously with other parameters of a person's behavior (ex: voice placement, movement in space, etc.), in order to provide more effective emotional categorization from postures.

## References

[1] N. H. Frijda, "The evolutionary emergence of what we call 'emotions,'" *Cogn. Emot.*, vol. 30, no. 4, pp. 609–620, May 2016.

[2] S. C. Soares, R. S. Maior, L. A. Isbell, C. Tomaz, and H. Nishijo, "Fast Detector/First Responder: Interactions between the Superior Colliculus-Pulvinar Pathway and Stimuli Relevant to Primates," *Front. Neurosci.*, vol. 11, p. 67, Feb. 2017.

[3] B. de Gelder, "Towards the neurobiology of emotional body language," *Nat. Rev. Neurosci.*, vol. 7, no. 3, pp. 242–249, Mar. 2006.

[4] "Body Language (1920×1080)." [Online]. Available: https://www.kcbi.org/wp-content/uploads/2017/07/body-language.jpg. [Retrieved: Apr, 2018].

[5] "Body language | Wondersbook." [Online]. Available: https://wondersbook.wordpress.com/2014/04/29/body-language/. [Retrieved: Apr, 2018].

[6] 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.

[7] "CMU-Perceptual-Computing-Lab:OpenPose."[Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/openpose. [Retrieved: Apr, 2018].

[8] "TensorFlow." [Online]. Available: https://www.tensorflow.org/. [Retrieved: Apr, 2018].

[9] A. C. Schapiro, T. T. Rogers, N. I. Cordova, N. B. Turk-Browne, and M. M. Botvinick, "Neural representations of events arise from temporal community structure," *Nat. Neurosci.*, vol. 16, no. 4, pp. 486–492, Apr. 2013.

[10] "Keras Documentation." [Online]. Available: https://keras.io/. [Retrieved: Apr, 2018].

[11] "Keras: Deep Learning for humans."[Online]. Available: https://github.com/keras-team/keras. [Retrieved: Apr, 2018].

[12] N. Pittaras, F. Markatopoulou, V. Mezaris, and I. Patras, "Comparison of Fine-Tuning and Extension Strategies for Deep Convolutional Neural Networks," vol. 10132 LNCS, 2017, pp. 102–114.