Analysis of Emotions from Body Postures 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* — In this paper, we present a state of the art regarding computer vision systems for the detection and classification of human body postures. Although emotions conveyed by human body postures are an important means of socio-communicative functions in several contexts, a surprising lack of systems that enable the recognition and classification of postures for different emotional signatures, is acknowledged. The neglect of such systems is most likely due to the complexity of the emotions reflected in body postures, and to the wide range of variations in postures, particularly when assessing groups of individuals. Despite the existence of several sensors, allowing to obtain images of various types, from color images to thermal images, no one was yet used for this purpose. We propose the use of a recently developed algorithm, which has presented optimal results in several domains, to allow for the recognition of emotions from human body postures.**

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

I. INTRODUCTION

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 [1], particularly in the fields of cognitive, affective and social neuroscience [2][3]. 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 [4], to teamwork [5] and mental health contexts (e.g., postures associated with psychopathology, such as unipolar depression [6]).

Our proposal is that, by using digital cameras and algorithms that allow to extract human body postures, postures associated with different emotional dimensions should be mapped. We suggest the development of a system that allows the detection and classification of these postures, in groups of individuals, which are asked to freely interact in a dynamic. 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) [7][8][9]. This raises the following questions: "How to use a PC and a camera to measure 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 1 gives a brief introduction and presentation of problem, in Section 2 are presented the existing computer systems for the presented problem, Section 3 discusses the definition of digital image, presenting the various types of existing image sensors, Section 4 are addressed some of the existing posture detection algorithms and in section 5 a conclusion is made about everything that has been said previously.

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 [10], 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

 ISBN: 978-1-61208-638-5

relationship between the movements of the students and their attention.

In [11], a system for the recognition of human actions based on posture primitives is described. This system, like [12], 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 [13].

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 shows an array of pixels relating to a digital image and Figure 2 represents the gradient of a Red, Green and Blue (RGB) image by merging pixels.

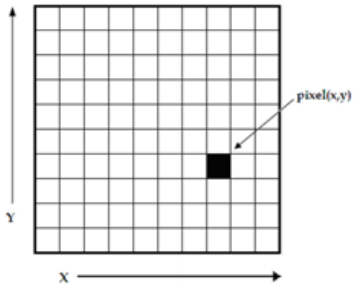

Figure 1. Representation of an array of pixels in an Image with Width X and Height Y [14].

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 3 shows an intensity matrix of a grayscale image for a given area [16].

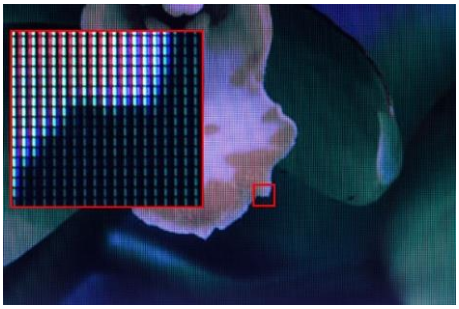

Figure 2. Gradient associated with a region of an RGB image [15].

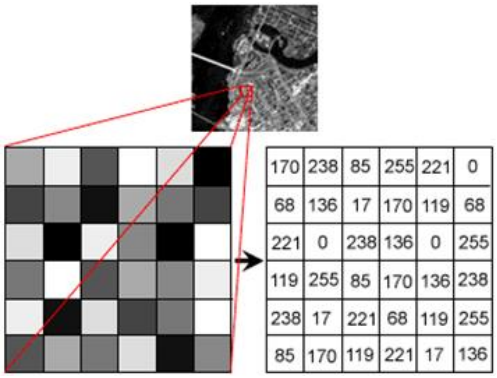

Figure 3. Matrix for certain area of a grayscale image [17].

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 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 [18].

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 4 shows the operation of this type of filter.

 ISBN: 978-1-61208-638-5

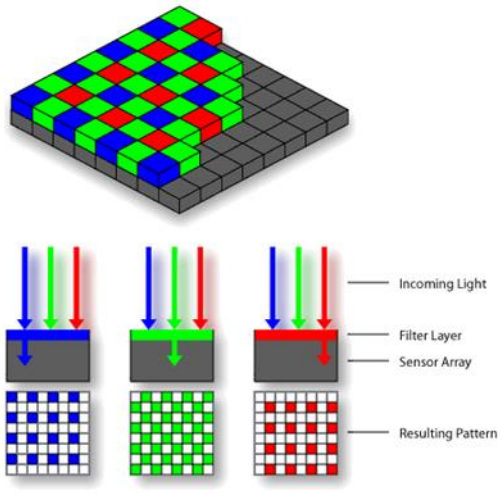

Figure 4. Application of a Bayer filter to obtain a color image [19].

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

Figure 5 shows an example of a CCD and CMOS sensor.

Figure 5. Example of CCD (left) and CMOS (right) sensor [20].

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 [21]. 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 6 shows a thermal image, obtained through a *Flir* [22] 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.

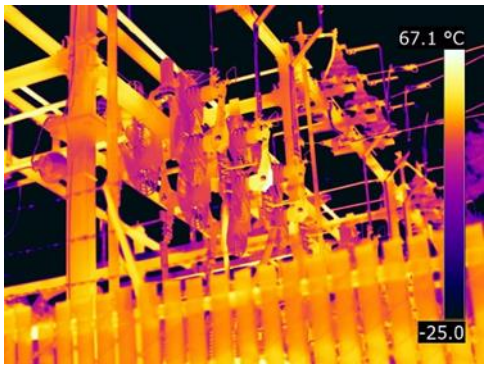

Figure 6. Example of CCD (left) and CMOS (right) sensor [22] .

This type of sensor can be used even in low-light environments, as opposed to sensors, such as CCD and CMOS [21]. There are several areas where these apply. From security, where they can be used to detect intruders even in low light situations [23], to the industry, where they can be used to detect heating problems in machines, which are not detected by the human eye [21], passing through the detection of people through the temperature of the human body [24].

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 [25]. 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 [26].

Figure 7 shows the comparison between multispectral and hyperspectral images.

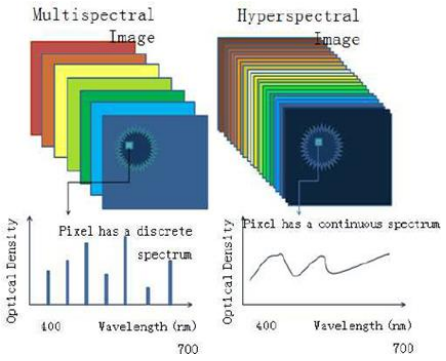

Figure 7. Comparison between a multispectral (left) and hyperspectral (right) image [27].

 ISBN: 978-1-61208-638-5

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 [28]. 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 [29]. 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 [30].

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 [9], or simply through the previous teaching of the system for the intended postures [8].

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. [8][31] and in situations that involving groups of people, where some parts of the body overlap [31]. 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 [32]. Thus, it is possible to obtain, with only one device, different types of images. Figure 8 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 9 shows the steps explained above.

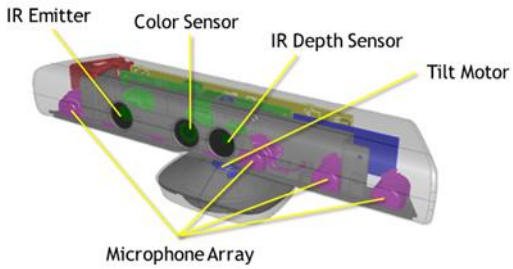

Figure 8. Hardware Configuration of a Kinect Device [33].

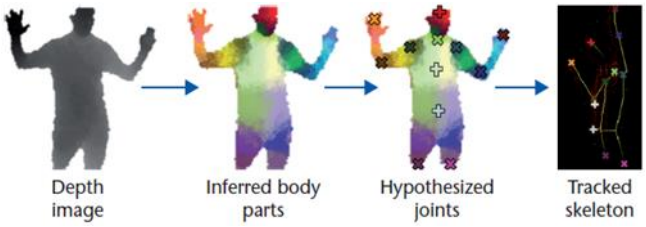

Figure 9. Detection steps of the Human Skeleton through the Kinect Skeletal Tracker Software [34].

In April 2017, the OpenPose library [35] was launched. Using only RGB images, this library can detect and extract 2D values from the main parts of the human body. In 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 [31], where a neural network is used to simultaneously predict confidence maps for body part detection (Figure 10b) and affinity fields for association of parts of the body (see Figure 10c), 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 10d). 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 10e).

In [36][37], are presented approaches of detection multiple human skeletons in simple RGB images with efficient results, however fall short of [31].

 ISBN: 978-1-61208-638-5

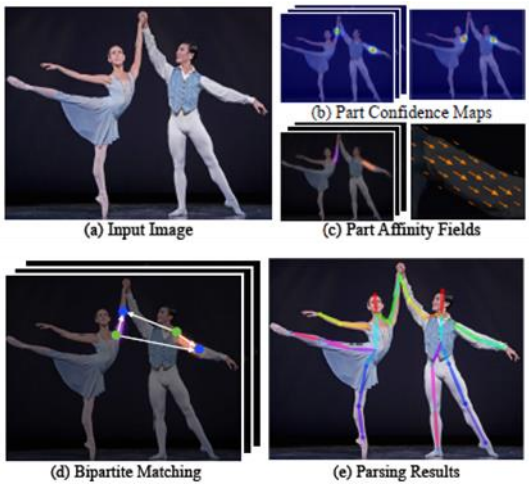

Figure 10. Detection of the Human Skeleton through the OpenPose library [31].

V. 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.

Finally, the Human Posture Detection algorithms research work [31] 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.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] E. Babad, "Teaching and Nonverbal Behavior in the Classroom," in *International Handbook of Research on Teachers and Teaching*, Boston, MA: Springer US, 2009, pp. 817–827.

[5] 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..

[6] 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. 40–9, Aug. 2013.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] T. 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.

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

[14] 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.

[15] "Free How to Photoshop Tutorials, Videos & Lessons to learn Photoshop training | Photoshop Course." [Online]. Available: http://www.we-r-here.com/ps/tutorials/. [Retrived: Apr, 2018].

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

[17] "Naushadsblog." [Online]. Available: https://naushadsblog.wordpress.com/. [Retrived: Apr, 2018].

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

[19] "Wikimedia Commons." [Online]. Available: https://commons.wikimedia.org/wiki/Main_Page. [Retrived: Apr, 2018].

[20] "Photography tips and tricks, Equipment, Photography News, Photography Books, Tutorial, and Lighting - OneSlidePhotography.com." [Online]. Available: http://oneslidephotography.com/. [Retrived: Apr, 2018].

[21] 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.

[22] "FLIR Systems | Thermal Imaging, Night Vision and Infrared Camera Systems." [Online]. Available: http://www.flir.eu/home/. [Retrived: Apr, 2018].

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

[24] 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.

[25] L.-J. Ferrato and K. W. Forsythe, "Comparing Hyperspectral and Multispectral Imagery for Land Classification of the

 ISBN: 978-1-61208-638-5

Lower Don River, Toronto," *J. Geogr. Geol.*, vol. 5, no. 1, pp. 92–107, 2013.

[26] "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. [Retrived: Apr, 2018].

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] J. Han, L. Shao, D. 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.

[33] "Kinect for Windows Sensor Components and Specifications." [Online]. Available: https://msdn.microsoft.com/en-us/library/jj131033.aspx. [Retrived: Apr, 2018].

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

[35] "OpenPose - Realtime Multiperson 2D Keypoint Detection from Video | Flintbox." [Online]. Available: https://cmu.flintbox.com/public/project/47343/. [Retrived: Apr, 2018].

[36] E. Insafutdinov, M. Andriluka, L. Pishchulin, S. Tang, E. Levinkov, B. Andres et al., "ArtTrack: Articulated Multi-person Tracking in the Wild," Dec. 2016.

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

 ISBN: 978-1-61208-638-5