Design of Algorithms for People Segmentation Using in Scenes for Non-Invasive Monitoring of Vital Signs Through Video

Gabriel Brenes-Vega, Oscar González-Alfaro and Luis Chavarría-Zamora

May 22, 2022

Design of algorithms for people segmentation using in scenes for non-invasive monitoring of vital signs through video

Gabriel Brenes-Vega, Oscar González-Alfaro, Luis Chavarría-Zamora
emails: gebv0896@gmail.com, oscar080899@gmail.com, lachavarria@tec.ac.cr
Computer Engineering Department
Instituto Tecnológico de Costa Rica
Luis Chavarría-Zamora

***Abstract*—This project is the result of multiple investigations related to technologies that allow manipulation of the elements present on stage. These investigations provided a broader overview regarding the possible use of technologies and allowed selecting the tools that provide greater benefit and adaptability to future changes in the project. This implied that a new system was developed, which allows detecting the faces present in a video from a file or a video in real time and later segmenting and saving each face detected in an individual video, for this two algorithms are used, one of them implementing an artificial intelligence approach and the other a traditional image processing approach in order to compare them at the end of development and thus determine which of the two is more optimal in terms of use of computational resources for the execution of the system in question. Each one is analyzed taking into account the least amount of resources used, to later extract the vital signs per person using Eulerian Magnification. With this project, vital signs monitoring systems will benefit, since large and complex instruments that must be connected to the body of patients that may bother them or limit their care will no longer be required, since now the entire monitoring process of the vital signs will be performed only by analyzing the images of a video using the Euler Magnification algorithm, in this way a non-invasive monitoring is guaranteed that avoids having to connect devices to the patient's body.**

***Index Terms*—Traditional image processing, artificial intelligence, vital signs, Eulerian motion magnification, face detection, segmentation.**

I. INTRODUCTION

Currently, the devices used in health centers for monitoring the vital signs of patients consist of large and complex machines that generally must be connected to the patient by means of instruments such as cables, tubes, injections, patches and others, which sometimes, can be uncomfortable and hinder the tasks that health personnel must carry out for patient care [1]. If the patient is in a very delicate situation, the situation can be much more complicated because the manipulation of the patient with so many devices connected to his body [1] and this together with his already delicate state of health greatly increases the complexity of his care.

In summary, when it comes to looking after the life of people in critical conditions, most of the time instruments are used that are not comfortable for the patient, which greatly hinders their care and can even cause minor injuries or in certain cases [1], considerable damage which leads to an important problem to deal with. Therefore, the objective of the product being developed is to simplify the conditions of current vital signs monitoring equipment and greatly facilitate the medical care of patients, since the multiple devices that are placed on the body of the patients will be dispensed with and in their place, it will give way to a monitoring by means of video which does not invade the space of the person. Using video formats is a great advance in this field, since there is certain information that may be imperceptible to the human eye, but for elements such as a video camera, these are more accessible. This advance, together with the increase in the production of smartphones, may imply a new path that allows the development of new applications for the processing of video formats in the area of medicine [2]. Returning to the subject of monitoring vital signs, it can be mentioned that it is a field of medicine that has been widely addressed by various entities such as in [3] [4]. Following the idea of non-invasive systems, our project has the objective to take a video in real time and detect each face present in said video to later segment each of these faces that were detected in individual videos with the aim that in future implementations, it is possible to take each one of these videos to analyze them according to the data that you want to obtain.

The face detection process and the subsequent segmentation of the same in individual videos was carried out using two implementations, one of them using the traditional digital image processing approach and the other using a machine learning approach, in the end this project makes a comparison between the two implementations to determine which was more efficient in terms of use of computational resources. The proposed system can become a great ally to the Eulerian Magnification algorithm, which can have the ability to be used to extract the vital signs of a person, as explained in [2], in addition to providing a step to the scaling of these systems. Next, a series of antecedents are presented, which will allow establishing the bases as well as a context to better understand the purpose of the project. Following is the description section of the developed solution, in which the implementations that were developed are explained in greater detail. This section includes a series of diagrams that expose the workflow that was chosen, in addition to mentioning the tools and modules necessary for its development. Subsequently, the results obtained from both developed methods are presented, as well as a comparison between them. After this section we can find the part of conclusions and future work, which focuses on formally presenting the analysis of the results and giving the reader a preview of the impact that this research may have on future work. Finally, there is the references section, which contains all the documents that were taken as reference for the preparation of this project.

II. BACKGROUND

The vital signs monitoring study has been the inspiration for a large amount of research worldwide, which aims to improve the stay of patients in medical centers. The Salesian Polytechnic University was not far behind and began with the design and development of a prototype for non-invasive vital signs monitoring [5], which implements data communication to mobile devices. Its main objective is to minimize discomfort and avoid putting patients at risk when they are required to capture vital signs. Regarding the system [5], it can be mentioned that it captures a series of status readings from different sensors present on a programmable card *ChipKIT*. Going into detail, we can mention the implementation of *serial*, *I2C* and *SPI* communication protocols, both to present and transmit the data through the *Bluetooth* communication protocol. Subsequently, a signal processing is applied, and the resulting values are sent to the mobile device, so that the user can observe the condition in which it is in a more friendly way.

Philips is one of the companies that has started to develop the idea of non-invasively monitoring vital signs through video recording. To date, the system [6] uses non-contact video technology to detect vital signs, which has already been endorsed by a scientific study in which 41 people participated. Levels of oxygen saturation in the blood, heart rate and respiratory rate are the signs that the system is able to pick up. To go into detail, it can be mentioned that the technology is based on the detection of slight changes in the color of the skin of the patient's forehead, which cannot be perceived by the human eye. Once these changes are detected, an algorithm is used, which has the purpose of using the information pertinent to the color transitions of the person's forehead, to measure the oxygenation level of the arterial blood. Philips focuses this new technology on those premature or newborn babies, since currently the methods used in the aforementioned cases are carried out by means of a series of sensors that adhere to the body, which can cause damage to the skin of babies.

The world demand for smartphones this year has increased by approximately 7% [7] compared to previous years. This factor was taken into consideration by the research group *TELETECNO* of the Francisco José de Caldas District University. The institution has carried out research and proposals focused on the implementation of control electronics and telecommunications technologies in order to monitor people's vital signs. The main focus of the institution is to provide new tools to those people who suffer from chronic diseases and who require constant supervision of the state of health. Faced with such a situation, the research group seeks to take advantage of the benefits offered by new technologies to develop an electronic health system platform that can complement the health services offered by medical centers. Regarding the project, it can be mentioned that the group proposes a system for monitoring and recording the bodily signs of people [8], in this way, in the case of the values that are recorded are above the thresholds of normal levels, a notification is sent to the person to seek medical attention. The operation of the system is based on the revolution pertaining to *Tics* as well as the convergence of devices that present the possibility of communications through *WAN* and *LAN* networks, in this way it is intended develop a safe non-invasive environment for the person.
One of the main antecedents and one of the studies that functioned as the basis on which this project was raised is the Eulerian movement magnification algorithm [9], which was developed by researchers at the Massachusetts Institute of Technology in 2012. This algorithm makes it possible to amplify the movement present in the videos and that is imperceptible or very little perceptible to the human eye, for this the spatial decomposition is applied, followed by a temporal filtering to the *frames* of the video. With the use of this algorithm it is possible to detect, for example, the flow of blood through the face or the slight movement that is generated in the chest caused by breathing.
Additionally, it is important to mention the work exposed in [2] in which a more detailed analysis of the Euler Magnification algorithm is made, treating the subject from a point of view more of optimization of the procedure in terms of consumption of resources, it should be noted that what is shown in the aforementioned document constitutes the next step for the project that is presented in this work; another source of great value for this topic of vital signs monitoring is the work presented in [10] in which a radar is used to detect frequency waves caused by disturbances generated by respiration.
As for more daily use applications, the case of the mobile application called Binah.ai can be mentioned [11], this allows to determine aspects such as heart rate, respiratory rate, oxygen saturation in the blood, stress level among others; just by analyzing people's faces through the cell phone camera.

III. Description of the developed solution

Currently, the vast majority of vital signs monitoring devices for patients are individual in nature, that is, a single device is designed to watch over a single patient in addition to the large number of instruments that must be placed on it. Therefore, it is proposed to implement a non-invasive vital signs monitoring system through video, which can simultaneously analyze each person that appears in said video and determine for each of them their heart and respiratory rates, to achieve the previously described, people segmentation techniques must be used to detect them in the recording. For the development of this project, the main objective is to achieve segmentation using two techniques, one of traditional image processing and the other using artificial intelligence and finally make a performance comparison between them and determine which of the two is more optimal to carry out this work taking as a point of reference the consumption of computational resources.

Before going into detail for both the image processing part and the artificial intelligence part, it is necessary to keep in mind the general context of the system. Basically, the algorithm will receive a video transmission in *MP4* format or a video transmission in real time from a camera, for example, to which the detection and segmentation methods of people developed with the previously mentioned techniques. Later, the system will return individual videos corresponding to the segmentation of the individuals. To clarify the aforementioned, in Fig. 1 the general operation of the system is appreciated.

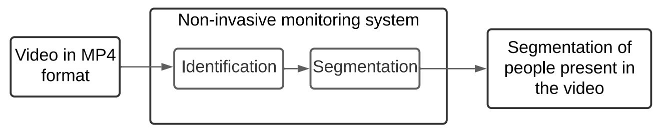

Fig. 1. Second Level Diagram of the implemented system

Each of the implemented algorithms based on the selected approaches is explained in more detail below:

A. Traditional Image Processing

When talking about traditional image processing, reference is made to all those techniques and procedures that allow images to be manipulated to transform them or obtain data from them, but which, as detailed in [12], have certain limitations in their sensitivity. analysis of external disturbances that may modify the images, precisely for this reason artificial intelligence methods have been developed that allow optimizing the processes on the images and better address these limitations.
The traditional image processing algorithm that was implemented for this project was developed in the programming language *Python* with the help of the computer vision library *OpenCV* and the *psutil* library for data capture consumption of resources such as CPU, RAM and others; this algorithm has the stages described in Fig. 1. The tools and techniques used in each of these stages are described below:

1) Video capture from the respective source: In this phase, the video to be processed is captured, either by means of a video in *MP4* format or a video from a real-time transmission such as using a camera, in the case of this project, the computer's own camera which has 0.92MP (*MegaPixels*) and a maximum resolution of 1280x720. To capture this video, the *VideoCapture* method of the *OpenCV* library is used, with this method you can obtain the video either from a file or from a source, in this case, the computer camera.

2) Face detection: This stage is in charge of analyzing the video to check if there are faces present. For this work a function called *CascadeClassifier* from the *OpenCV* library was used which uses a *XML* file that contains a pre-trained model with less than 15 training iterations and which is already provided in the *OpenCV* documentation, this file was pre-trained with Haar's cascades method. The Haar waterfall method was proposed by Paul Viola and Michael Jones in 2001, it is based on a series of “Haar characteristics” as it is appreciated in Fig.3, that consist of a series of rectangles with black and white areas that pass through the entire image or in this case, for all the *frames*, these allow identifying characteristics that help to identify that what is being analyzed is a face or not by detecting areas darker than others, such as the eye area with regarding the cheeks, the area of the sides of the nose with respect to the area of the nasal septum, among others; to improve the detection process, a series of cascade classifiers are used, each one is in charge of discarding characteristics that do not correspond with a face, as each image passes through one of these classifiers it is ensured that it represents a face, to speed up the entire process mentioned above, a technique called *AdaBoost* is used. The events described are part of the work referred to in [13] and the algorithm implemented in the "traditional image processing" part of this project for face detection is based on [14].

3) Segmentation of each face detected: In order to segment each of the faces that were detected, the *CentroidTracker* class is used, which was implemented in [15], this allows tracking of each of the faces detected by the screen and thus identify in every moment that a face is the same as the previous *frame*, in case the face moves from one position to another relatively close from one moment to another or disappears for a time no longer than 2 seconds and then returns to appear in the video, the *Centroid Tracker* will be able to identify this face as itself and not as another face. To view the detected faces, this *CentroidTracker* class stores the coordinates of each face and from these draws a red box that frames it and at the same time indicates a unique identifier for each of them.

4) Saving each segmented face into an individual video: With the help of the *CentroidTracker* it is possible to identify each face as the same from one *frame* to another and thus be able to save each capture in the corresponding video file. In the Fig.2 there is a diagram showing the working flow of the complete algorithm that was developed under the traditional image processing approach:

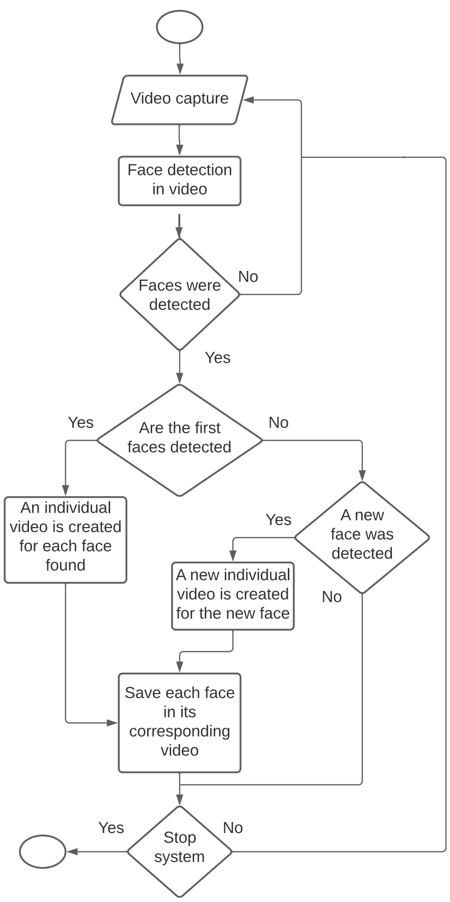

Fig. 2. System process flow implemented with traditional image processing

Fig. 3. Haar Characteristics

B. Artificial intelligence

Before beginning to detail the developed algorithm in depth, it is necessary to know a series of concepts to provide a context of the situation. The first term to take into consideration is artificial intelligence, which refers to those systems that mimic human intelligence to carry out tasks, in addition to having the characteristic that they can improve their performance interactively based on the information that collect. At present, artificial intelligence or better known as *AI* for its acronym, has become one of the most widely implemented technologies in applications that require the completion of complex tasks that previously required human input. This technology has a number of subareas, but the most popular on the market today is known as *machine learning*. It is worth mentioning that the implementation of any of the branches is determined based on the purpose of the system to be developed. For example, when *machine learning* is integrated, the system [16] is expected to have the ability to improve its performance relative to the data it consumes. There are a series of steps [17] that it is recommended to follow when implementing *machine learning*, which are:

1) Understand the problem
2) Define evaluation criteria
3) Evaluation of the solution
4) Organize the data
5) Build the model
6) Metric Integration

Today there are a wide variety of open source libraries focused on implementing machine learning through tasks. One of them is known as *Tensor Flow*, which presents a focus on training a shared global model, which preserves the training data locally in the [18] system. One of the advantages offered by the library is that it has a series of pre-trained models in various fields, which allows the user to avoid the step corresponding to model training. Once the bases are already established in regards to artificial intelligence, we can continue to talk in greater detail about the implemented algorithm.

Regarding the developed system, it can be mentioned that it was made using the *Python* programming language, as well as a series of libraries that were installed through a package management system. The name and operation of the libraries used are detailed below:

- *OpenCV*: manipulation of multimedia files.
- *Face Recognition*: stores the pre-trained model in identifying people's faces.
- *psutil*: provides a series of metrics regarding the computational resources used.

The system work process begins with the capture of video in real time from the computer camera through the *OpenCV* library. Next, a *frame* is selected, to which a size reduction

is applied, in order to increase the processing speed. Subsequently, the model found in the *Face Recognition* library is instantiated, which, as previously mentioned, already has training for face recognition. In short, the model parses the *frame* in search of possible faces, once it finds them it checks if there is already a similarity in the list of encodings. If there is a similarity, the index corresponding to the face is stored in a variable, together with its location, if there is not, a coding of the face is stored with a respective index, which is unique for each individual and also saves the coordinates referring to the face. Subsequently, an escalation of both the *frame* and the locations of the faces is carried out, in order to integrate a frame that delimits the face of the individuals and also presents the index of each of them from a visual way. The aforementioned, is kept in constant execution until the moment the user decides to finish executing the system. Once it is finished, a folder is created, which stores a series of videos, categorized by the indexes corresponding to each of the segmented faces.
In order to better clarify this process, in Fig.4 shows the system state diagram.

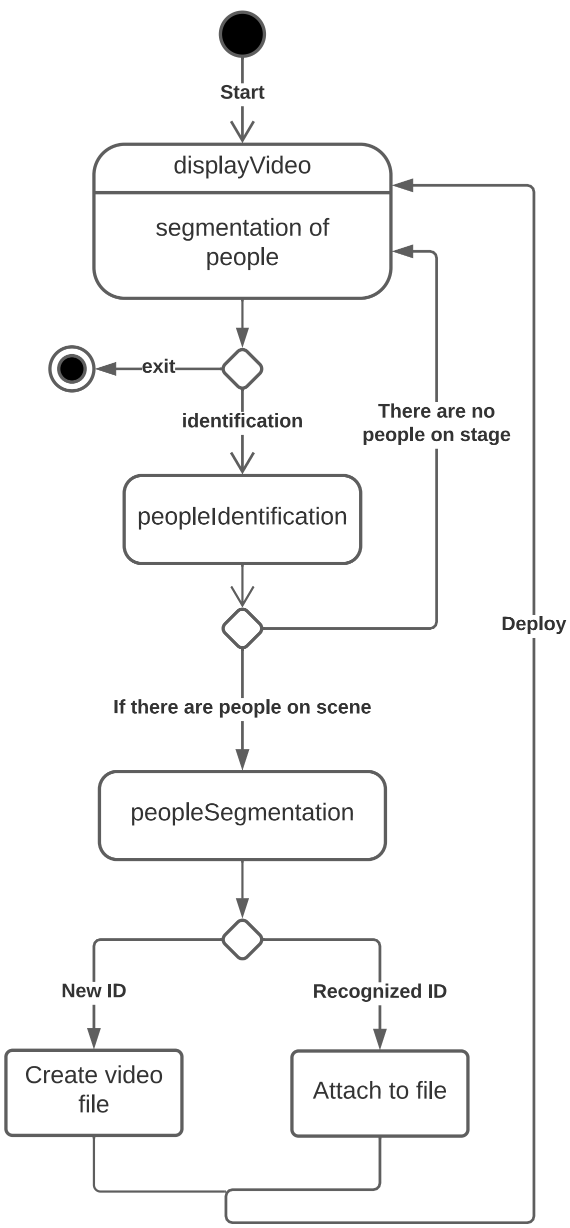

Fig. 4. System state diagram implemented with artificial intelligence

IV. Results obtained

A. Results obtained by the algorithm that integrates artificial intelligence

Regarding the algorithm that integrates the artificial intelligence technology, it can be mentioned that its functionality complies with the provisions of the requirements document. In Fig. 5 it can see the detection of a face present on the scene, to which the system assigned the identifier number zero. Regarding the segmentation part, in Fig. 6 the result of this process can be seen, that is, an individual video of the previously detected person.

Fig. 5. Face detected using the algorithm developed with artificial intelligence

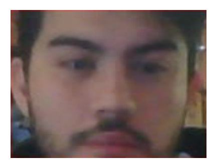

Fig. 6. Detected face segmentation using the algorithm developed with artificial intelligence

In order to know in detail the performance of the algorithm, the *psutil* library was integrated, which provides us with a series of values referring to the computational resources used by the system. The values presented below correspond to an average of 10 tests that were performed on the algorithm using the aforementioned library:

- CPU: 50.45
- RAM: 61.55
- Virtual Memory: 61.55
- Swap Memory: 82.38

As it is observed in the exposed values, the largest computational resource consumed is swap memory, on the contrary, the magnitude of the CPU is the resource with the least use.

B. Results obtained by the algorithm implemented with a traditional image processing approach

This algorithm satisfactorily meets the objectives established for this project, which basically consisted of being able to detect faces in a video and be able to segment each of them and store them in individual videos, these results are shown below:

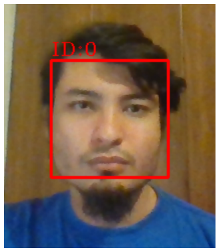

Fig. 7. Face detected using the algorithm developed with traditional image processing

Fig. 8. Detected face segmentation using the algorithm developed with traditional image processing

In Fig. 7 the detection of the face in the video made by the algorithm is shown and in Fig.8 the video generated and saved by the same algorithm for that face can be seen detected.

In order to know in detail the performance of the algorithm, the *psutil* library was integrated, which provides us with a series of values referring to the computational resources used by the system. The values presented below correspond to an average of 10 tests that were performed on the algorithm using the aforementioned library:

- CPU: 68.30
- RAM: 71.35
- Virtual Memory: 71.40
- Swap Memory: 82.48

As it is observed in the exposed values, the largest computational resource consumed is swap memory, on the contrary, the magnitude of the CPU is the resource with the least use.

The results shown above were obtained by executing both algorithms, both traditional image processing and artificial intelligence on a computer with Windows 10 Pro 64-bit, an Intel Core i5-7300 processor at 2.60GHz with 2 physical and 2 logical cores and an 8GB RAM memory, in addition, to obtain more objective results at the time of testing, the only user program running was the algorithm that you wanted to test running in Visual Studio Code.

C. Comparison

In Fig.9 the consumption of computational resources of both algorithms can be better appreciated.

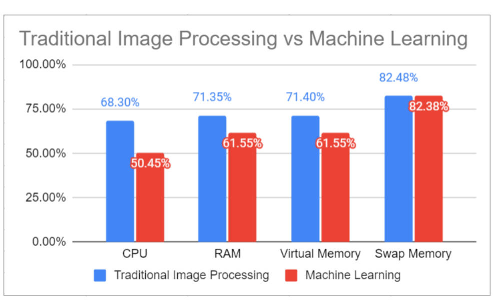

Fig. 9. Comparison between algorithms based on the consumption of computational resources

As it is observed in Fig.9, the algorithms show visible differences in terms of resource consumption. Regarding the CPU, a difference of 17.85% can be mentioned, where the image processing algorithm presents the highest value. In the rest of the data, such as RAM and virtual memory, the differences between the results are minimal, but there is a slight advantage on the part of the algorithm with the artificial intelligence approach. The aspect in which they present similar values is swap memory. Both oscillate between 82%, the artificial intelligence algorithm being slightly lower.

V. Conclusions and Future Work

In conclusion, two algorithms were implemented which satisfactorily fulfilled both the real-time video face detection tasks as well as the segmentation of each person detected in individual videos. Regarding metric terms, it can be mentioned that the algorithm implemented under the machine learning approach was more efficient in terms of the use of computational resources than the one implemented under the traditional image processing approach, in each of the four categories that were selected for this analysis, the process that uses an automatic learning approach resulted in favorable data with respect to traditional image processing where the CPU use is the main difference between machine learning approach and traditional image processing approach with an advantage of 17.85% over the traditional image processing, which established it as a very good option to be implemented in facial detection tasks in systems with limited resources, in the case of this work, the machine learning approach was more efficient since we use a highly pretrained and optimized model developed by the creators of the Labeled Faces in the Wild Home project [19] at the University of Massachusetts and its main feature is to reduce the computational resource consumption when carrying out facial detection tasks. In addition, given the extensive use that machine learning currently has, it has a large number of information sources and tools available. Regarding future work, thanks to this implementation, the Eulerian Video Magnification can be coupled with the work presented in [2]. Also it can be mentioned that integrating TinyML technology into the project will be analyzed in order to provide a certain level of adaptability in the event that it is required to integrate with embedded circuits. Likewise, the idea of further optimizing the code is proposed, to increase performance and significantly reduce latency.

References

[1] Hui, X., Kan, E. (2017). Monitoring vital signs over multiplexed radio by near-field coherent sensing. Nature Electronics, 1(1), 74-78. https://doi.org/10.1038/s41928-017-0001-0
[2] K. S. Lim, E. Moya-Bello and L. Chavarria-Zamora, "Resource Optimization of the Eulerian Video Magnification Algorithm Towards an Embedded Architecture," 2021 IEEE URUCON, 2021, pp. 576-579, doi: 10.1109/URUCON53396.2021.9647386.
[3] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, "Monitoring Vital Signs and Postures During Sleep Using WiFi Signals," IEEE Internet of Things Journal, vol. 5, no. 3, pp. 2071–2084, Jun. 2018, conference Name: IEEE Internet of Things Journal.
[4] F. Fioranelli, J. Le Kernec, and S. A. Shah, "Radar for Health Care: Recognizing Human Activities and Monitoring Vital Signs," IEEE Potentials, vol. 38, no. 4, pp. 16–23, Jul. 2019, conference Name: IEEE Potentials
[5] E. Tintín Durán, "Diseño y elaboración de un Prototipo de monitor de signos vitales aplicando métodos no invasivos con comunicación de datos a dispositivos móviles", 2021. [Online]. Available: https://dspace.ups.edu.ec/bitstream/123456789/7982/1/UPS-CT004847.pdf
[6] "Diseñan un sistema de cámaras que monitoriza los signos vitales sólo con ver al paciente", Consalud, 2021. [Online]. Available: https://www.consalud.es/tecnologia/disenan-un-sistema-de-camaras-que-monitoriza-los-signos-vitales-solo-\con-ver-al-paciente_27455_102.html
[7] "Venta mundial de smartphones se recupera con fuerte repunte de marca china", Larepublica.net, 2021. [Online]. Available: https://www.larepublica.net/noticia/venta-mundial-de-smartphones-se-recupera-con\-fuerte-repunte-de-marcha-china#:~:text=La%20venta%20mundial%20de%20smartphones,repunte%20de%20la%20marca%20Xiaomi
[8] O. Sarmiento Goméz and J. Rubio Cristiano, "Monitoreo remoto de signos corporales y transmisión de datos y alertas a una aplicación instalada en un smartphone", Repository.udistrital.edu.co, 2021. [Online]. Available: https://repository.udistrital.edu.co/bitstream/handle/11349/13383/SarmientoG%C3%B3mezOscar2018.pdf?sequence=2&isAllowed=y.

[9] Wu, Hao-Yu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. "Eulerian Video Magnification for Revealing Subtle Changes in the World." ACM Transactions on Graphics 31, no. 4 (1st of July, 2012): 1–8.

[10] F. Fioranelli, J. Le Kernec, and S. A. Shah, "Radar for Health Care: Recognizing Human Activities and Monitoring Vital Signs," IEEE Potentials, vol. 38, no. 4, pp. 16–23, Jul. 2019, conference Name: IEEE Potentials.

[11] "Video-based Vital Signs Monitoring - Binah", Binah, 2021. [Online]. Available: https://www.binah.ai/es/

[12] O. Reyes Ortiz, M. Mejia and J. Useche Castelblanco, "Técnicas de inteligencia artificial utilizadas en el procesamiento de imágenes y su aplicación en el análisis de pavimentos", Revista EIA, vol. 16, no. 31, pp. 189-207, 2019. Available: https://revistas.eia.edu.co/index.php/reveia/article/view/1215/1229

[13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001, pp. I-I,doi: 10.1109/CVPR.2001.990517

[14] "OpenCV usa Cascadas Haar para la detección de rostros - programador clic", Programmerclick.com, 2021. [Online]. Available: https://programmerclick.com/article/1274494374/

[15] "Simple object tracking with OpenCV", Gist, 2019. [Online]. Available: https://gist.github.com/adioshun/779738c3e28151ffbb9dc7d2b13c2c0a

[16] "¿Qué es la inteligencia artificial (IA)?", Oracle.com, 2021. [Online]. Available: https://www.oracle.com/mx/artificial-intelligence/what-is-ai/

[17] J. Heras, "Las 7 Fases del Proceso de Machine Learning - IArtificial.net", IArtificial.net, 2021. [Online]. Available: https://www.iartificial.net/fases-del-proceso-de-machine-learning/

[18] "TensorFlow Federated", TensorFlow, 2021. [Online]. Available: https://www.tensorflow.org/federated.

[19] Rim, D., Kamrul Hasan, M., Puech, F., Pal, C. (2015). Learning from weakly labeled faces and video in the wild. Pattern Recognition, 48(3), 759-771. https://doi.org/10.1016/j.patcog.2014.09.016