Background subtraction and yolo algorithm: two methods for the detection of people in uncontrolled environments

Substracción de fondo y algoritmo yolo: dos métodos para la detección de personas en entornos descontrolados

Subtração de antecedentes e algoritmo de yolo: dois métodos para a detecção de pessoas em ambientes não controlados

Carlos Vicente Niño Rondón[1]
Sergio Alexander Castro Casadiego[2]
Byron Medina Delgado[3]

Received: September 05th, 2020
Accepted: November 10th, 2020
Available: January 12th, 2021

How to cite this article:
C. V. Niño Rondón, S. A. Castro Casadiego, B. Medina Delgado. "Background Subtraction and Yolo Algorithm: Two Methods for The Detection of People in Uncontrolled Environments," *Revista Ingeniería Solidaria*, vol. 17, no. 1, 2021. doi: https://doi.org/10.16925/2357-6014.2021.01.06

Research article. https://doi.org/10.16925/2357-6014.2021.01.06

1 Universidad Francisco de Paula Santander, Facultad de Ingeniería
Email: carlosvicentenr@ufps.edu.co
ORCID: https://orcid.org/0000-0002-3781-4564
CvLAC: https://scienti.minciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0001703664

2 Universidad Francisco de Paula Santander, Facultad de Ingeniería
Email: sergio.castroc@ufps.edu.co
ORCID: https://orcid.org/0000-0003-0962-9916
CvLAC: https://scienti.minciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0001132598

3 Universidad Francisco de Paula Santander, Facultad de Ingeniería
Email: @ufps.edu.co
ORCID: https://orcid.org/0000-0003-0754-8629
CvLAC: http://scienti.colciencias.gov.co:8081/cvlac/visualizador/generarCurriculoCv.do?-cod_rh=0000290157

Abstract

Introduction: This article is the result of research entitled “Signal processing system for the detection of people in agglomerations in areas of public space in the city of Cúcuta”, developed at the Universidad Francisco de Paula Santander in 2020.

Problem: The high percentage of false positives and false negatives in people detection processes makes decision making in video surveillance, tracking and tracing applications complex.

Objective: To determine which technique for the detection of people presents better results in terms of response time and detection hits.

Methodology: Two techniques for the detection of people in uncontrolled environments are validated in Python with videos taken inside the Universidad Francisco de Paula Santander: Background subtraction and the YOLO algorithm.

Results: With the background subtraction technique, we obtained a hit rate of 84.07 % and an average response time of 0.815 seconds. Likewise, with the YOLO algorithm the hit rate and average response time are 90% and 4.59 seconds respectively.

Conclusion: It is possible to infer the use of the background subtraction technique in hardware tools such as the Pi 3B+ Raspberry board for processes in which the analysis of information in real time is prioritized, while the YOLO algorithm presents the characteristics required in the processes in which the information is analyzed after the acquisition of the image.

Originality: Through this research, aspects required for the real-time analysis of information obtained in processes of people detection in uncontrolled environments were analyzed.

Limitations: The analyzed videos were taken only at the Universidad Francisco de Paula Santander. Also, the Raspberry Pi 3B+ board overheats when processing the video images, due to the full resource requirement of the device.

Keywords: Hit rate, Comparative, People detection, Background subtraction, YOLO algorithm.

Resumen

Introducción: Este artículo es resultado de la investigación titulada “Sistema de procesamiento de señales para la detección de personas en aglomeraciones en zonas de espacio público de la ciudad de Cúcuta”, desarrollada en la Universidad Francisco de Paula Santander en el año 2020.

Problema: El alto porcentaje de falsos positivos y falsos negativos en los procesos de detección de personas hace que la toma de decisiones en las aplicaciones de videovigilancia, seguimiento y localización sea compleja.

Objetivo: Determinar qué técnica de detección de personas presenta mejores resultados en cuanto a tiempo de respuesta y aciertos en la detección.

*Metodología***:** Dos técnicas para la detección de personas en entornos no controlados son validadas en Python con videos tomados dentro de la Universidad Francisco de Paula Santander: la sustracción de fondo y el algoritmo YOLO.

Resultados: Con la técnica de sustracción de fondo se obtuvo una tasa de acierto del 84,07 % y un tiempo de respuesta medio de 0,815 segundos. Asimismo, con el algoritmo YOLO, la tasa de acierto y el tiempo de respuesta promedio son del 90% y 4,59 segundos respectivamente.

Conclusión: Es posible inferir el uso de la técnica de sustracción de fondo en herramientas de hardware como la placa Raspberry Pi 3B+ para procesos en los que se prioriza el análisis de la información en tiempo real, mientras que el algoritmo YOLO presenta las características requeridas en los procesos en los que se analiza la información después de la adquisición de la imagen.

Originalidad: A través de esta investigación se analizaron los aspectos necesarios para el análisis en tiempo real de la información obtenida en los procesos de detección de personas en ambientes no controlados.

Limitaciones: Los videos analizados fueron tomados sólo en la Universidad Francisco de Paula Santander. Además, la placa Raspberry Pi 3B+ presenta efectos de sobrecalentamiento al procesar las imágenes de vídeo, debido a la gran cantidad de recursos que requiere el dispositivo.

Palabras clave: tasa de aciertos, comparativa, detección de personas, sustracción de fondo, algoritmo Yolo.

Resumo

Introdução: Este artigo é resultado da pesquisa intitulada "Sistema de processamento de sinais para detecção de pessoas em aglomerações em áreas do espaço público da cidade de Cúcuta", desenvolvida na Universidade Francisco de Paula Santander em 2020.

Problema: a alta porcentagem de falsos positivos e falsos negativos nos processos de detecção de pessoas torna complexa a tomada de decisões em aplicações de vigilância por vídeo, rastreamento e rastreamento.

Objetivo: Determinar qual técnica de detecção de pessoas apresenta melhores resultados em termos de tempo de resposta e acertos de detecção.

Metodologia: Duas técnicas de detecção de pessoas em ambientes não controlados são validadas em Python com vídeos feitos na Universidade Francisco de Paula Santander: Subtração de fundo e algoritmo YOLO.

Resultados: Com a técnica de subtração de fundo, obteve-se uma taxa de acerto de 84,07% e um tempo médio de resposta de 0,815 segundos. Da mesma forma, com o algoritmo YOLO, a taxa de acerto e o tempo médio de resposta são 90% e 4,59 segundos, respectivamente.

Conclusão: É possível inferir a utilização da técnica de subtração de background em ferramentas de hardware como a placa Pi 3B + Raspberry para processos em que a análise da informação em tempo real é priorizada, enquanto o algoritmo YOLO apresenta as características requeridas nos processos em qual a informação é analisada após a aquisição da imagem.

Originalidade: por meio desta pesquisa, foram analisados os aspectos necessários à análise em tempo real das informações obtidas nos processos de detecção de pessoas em ambientes não controlados.

Limitações: Os vídeos analisados foram realizados apenas na Universidade Francisco de Paula Santander. Além disso, a placa Raspberry Pi 3B + superaquece ao processar as imagens de vídeo, devido à necessidade de recursos completos do dispositivo.

Palavras-chave: Taxa de acerto, Comparativo, Detecção de pessoas, Subtração de fundo, algoritmo YOLO.

1. INTRODUCTION

The detection of people is one of the most interesting topics in research today. Several alternatives for detection processes have been proposed, such as computer vision techniques [1] and deep learning algorithms [2]. In uncontrolled environments, object detection becomes complex, and even in cases where the object of interest is the human body, the challenge is even greater [3], since in this type of process it is not possible to predict the behaviour of variables such as the number of people circulating simultaneously, their location with respect to the video camera and the luminosity level when the images are captured [4].

Background subtraction is a computer vision technique often used to separate moving objects from the image background [5]. In normal conditions, the hit rate when applying the background subtraction technique is around 90 %.[6], but being susceptible to variations in brightness and noise levels in the image, it is common that this value tends to decrease [7].

The YOLO (You Only Look Once) algorithm is presented as an alternative to the detection and classification processes, since through a single neural network it is capable of predicting the limits and probability of distinguishing an object as a person according to previous training, achieving a hit rate of around 90% [8]. The main deficiency in the processes implemented with the YOLO algorithm, lies in the errors of location of the detections [9].

This article presents the application and comparison of the background subtraction technique and the YOLO algorithm, according to the hit rate and the response time of the processing system, when analyzing video images captured in uncontrolled environments at the Francisco de Paula Santander University in Cúcuta, Colombia. The processing system consists of a 5 Mpx Raspberry Pi video camera, a Raspberry Pi 3B+ embedded board and programming is done in the Python programming language.

2. LITERATURE REVIEW

The following are high-impact investigations related to the application of the background subtraction technique and the YOLO algorithm focused on the people detection. This background refers to research processes developed since 2015.

For the background subtraction, Jeon *et al.* [6] proposes a system for the detection of people based on an infrared camera and generates the background of the images, to which a differentiation of pixels and edges is applied, in order to define the candidate regions for detection. By means of filtering, they remove the noise from the image and redefine the candidate regions according to the size information of the objects, considering the line of sight of the video capture device. Likewise, Zuo *et al.* [10] performs a detection based on improvements in the image background by modeling Gaussian mixtures. The method consists of 3 stages in which the image is segmented and the value of the pixels is replaced by an average value of pixels. Similarly, the technique of half-wave threshold is applied, accompanied by morphological operations for noise removal. Finally, the image background is updated in an adaptive way. In turn, Miranto, Sulistiyanti y Arinto [11] proposes a tracking system for detecting people based on adaptive background subtraction with Gaussian mix modelling, which adapts to dynamic changes in light levels, as well as to the movement of small objects that are ignored by the processing system. Liu *et al.* [12] performs a modeling process for the background subtraction of moving objects based on the characterization of components of the image background. The method was developed for grayscale images, although it can be replicated for color images. First, the components are initialized and updated in such a way that the information contained in them is potentialized. Then, a number of components are assigned and generated, thus allowing for the detection of moving objects in the foreground of the image. Also, Htun, Zin and Tin [13], are carrying out a study of image processing techniques for a surveillance system in the care of the elderly, in which they propose stages of object detection and feature extraction accompanied by a Markov model. Object detection is performed by means of the background subtraction technique in which they segment the background by means of a Gaussian mixture model, which facilitates the stages of feature extraction and decision, generating high hit rates. Likewise, Mariappan, Thong y Muthukaruppan [14] propose a methodology for the design of a video surveillance system for motion detection, based on the OpenCV library and a camera as a sensor. The detection algorithm consists of grayscale conversion stages, image background segmentation, smoothing filters and thresholding. The method is highly efficient and runs smoothly on both a computer and the BeagleBoard. Li *et al.* [15] develop a system for monitoring people with an aerial camera by means of algorithms based on neural networks, in which, as a first stage, an extraction of regions of interest is performed by subtracting the background, where they perform the subtraction of the object that circulates through the main frame with the background of the image, accompanied by filtering stages by morphological operations based on a 3x3 structural element and the operation of dilation and thresholding processes. In this way, the characteristics of interest are extracted and the decision process is improved. Also, Rantelobo *et al.* [16], developed a monitoring and people-counting system using an embedded Raspberry Pi 3 board, through image processing with web server communication. The processing consists of stages of conversion to grey, segmentation and subtraction of the image background, morphological operations and approximation of the image contour area. The detection system has high hit rates and a response time that theoretically guarantees the processing of information in real time.

As for the YOLO algorithm, Ren, Fang and Djahel [17], propose a method for detecting and counting people based on five stages; so that the video capture device is set first. In the second stage they detect people by additional training of the convolutional neural network, while in the third stage they count by means of a selection of limits. In the fourth stage, the information obtained in the count is analyzed while the last stage refers to the visualization of the detections. Likewise, Vaidya *et al.* [18] compare various configurations of neural networks and select a series of algorithms for the detection of people and compare them with the YOLO algorithm. They are able to process up to 45 frames per second, and for the YOLOv3 version, they demonstrate a high requirement of hardware tools, so they carry out the study with the Tiny YOLO version. In addition, Lucian *et al.* [19] propose a method for the detection of lower extremities of the human body, for which they perform a geometric approximation by processing images in 2D images in search of legs. Due to the variability of the input data, they use deep learning techniques such as the YOLO algorithm in version 2 and Tiny YOLO. By testing these techniques, it is obtained that the ideal method for training and evaluation is Tiny YOLO with batches of 64 images, gradient descent of 0.8 and learning rate of 0.01. Also, Ahmad, Imran and Adnan [20], propose a method for detecting and counting people from an aerial view using the YOLO algorithm. The dataset used for the training corresponds to frontal aerial images, while for the tests, they use the YOLO weight file in its version 3, which comes pre-trained. The location of the device that captures the images from a position that alludes to an aerial view, generates a reduction in the errors in the detections and allows a greater number of people to intervene during the processing of the images. Yang *et al.* [21] developed a system for adaptive object detection based on time differences. The detection is done through a platform called ToDo, focused on the tracking and tracing of objects, previously trained with the YOLO algorithm. The system presents hits of up to 93% and achieves a reduction of GPUs in the system of up to 17%. Likewise, Liu *et al.* [22] developed a system for the detection of attacks on people with sharp weapons based on the YOLO algorithm, as an alternative to conventional methods of metal and X-ray detection. Initially, the system had a low number of images for training and testing, which was increased with data amplification techniques. The system improves process detection by approximately 3 %. Also, Aktas *et al.* [23] propose a method for the design of assistance systems for the disabled by detecting surfaces when separating people from them using the YOLO algorithm with the training of approximately 4600 images. The success rate fluctuates between 70% and 90%. Bohush and Zakharava [24] developed a method for monitoring people indoors using convolutional neural networks using the YOLO algorithm is version 3. The color analysis is done in HSV, with a total of 29 convolutional layers and equals and improves upon conventional methods of detecting people. Real-world processing is achieved by using CUDA technology and NVIDIA GTX 1060 graphics card.

3. METHODOLOGY

In processes of people detection in uncontrolled environments, techniques and strategies are required to diminish processing errors due to overlapping, noise and variations in luminosity levels that affect process reliability [25]. A methodology based on three stages is proposed. The first stage concerns the location of the video capture device, taking into account factors such as the height and line of sight of the device. The second stage corresponds to the application of the background subtraction technique and the YOLO algorithm for the detection of people in uncontrolled environments. Finally, in the third stage, a comparison is made between the techniques applied with respect to the hit rate and the processing response time [26], as shown in Figure 1. The methodology developed for the background subtraction technique is approached from a mathematical point of view, because it is presented as a combination of basic concepts for image processing. Similarly, the methodology developed in the YOLO algorithm is approached from a computational point of view, because the algorithm is presented as an alternative for real-time object detection in computers without GPUs (graphics processing unit).

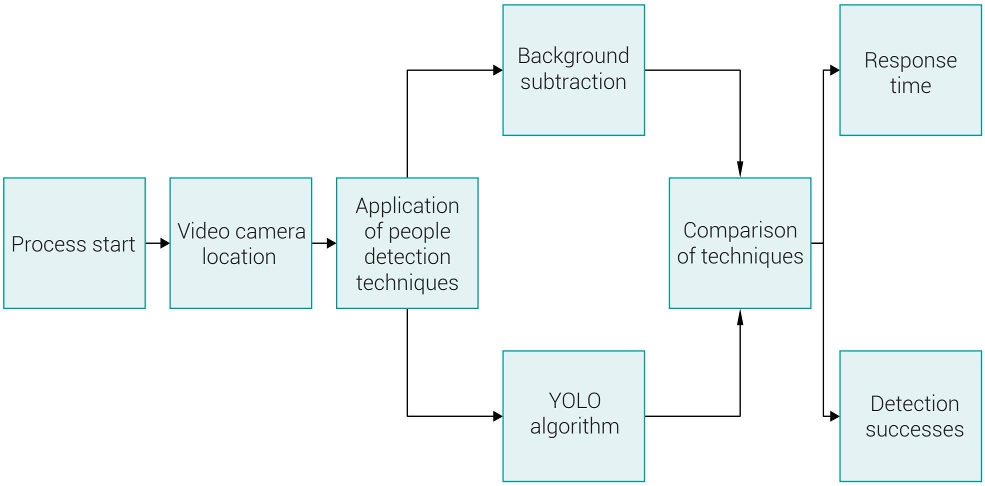

Figure 1. Proposed methodology.
Source: own work

Similarly, Figure 2 shows the hardware and software components of the processing system. The module consists of a 5 Mpx Raspberry camera connected via CSI standard to an embedded Raspberry Pi 3B+ board. The software tools used for the Python programming language and the specialized computer vision library OpenCV [27].

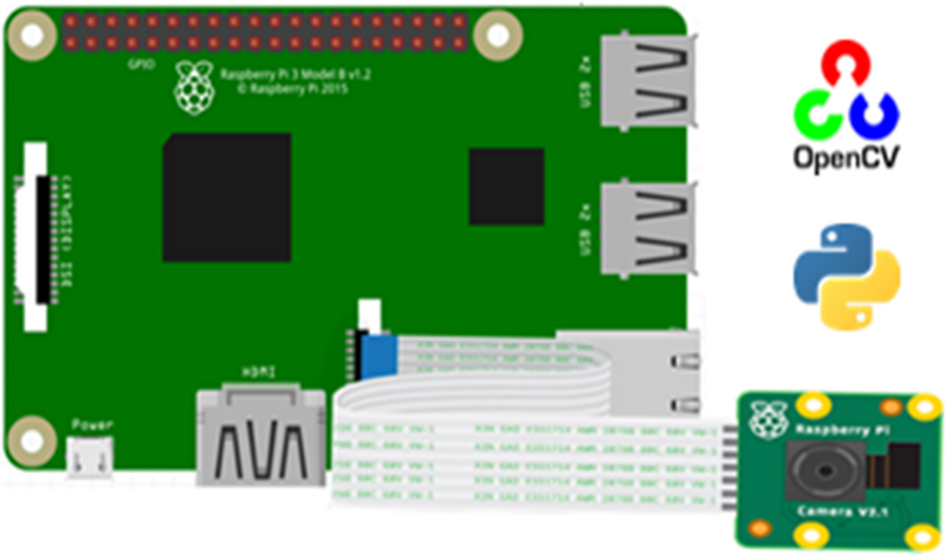

Figure 2. Processing system.
Source: own work

3.1 Video camera location

The device is located in the building Aula Sur SF at a height of 4.2 meters to avoid overlapping of people and have as many as possible in the visible range of the camera [28] [29]. A horizontal line of sight of up to 15 meters is presented. Figure 3 illustrates the video camera location inside the Francisco de Paula Santander University.

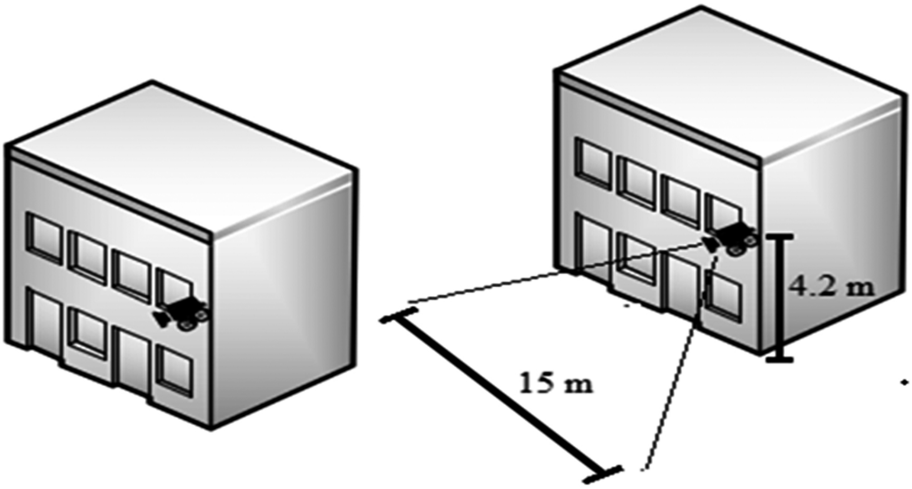

Figure 3. Video camera location.
Source: own work

3.2 Application of people detection techniques

In this stage, the necessary foundations for the application of the background subtraction technique and the YOLO algorithm are defined.

3.2.1 Background subtraction

Figure 4 shows the block diagram describing the process for people detection by background subtraction.

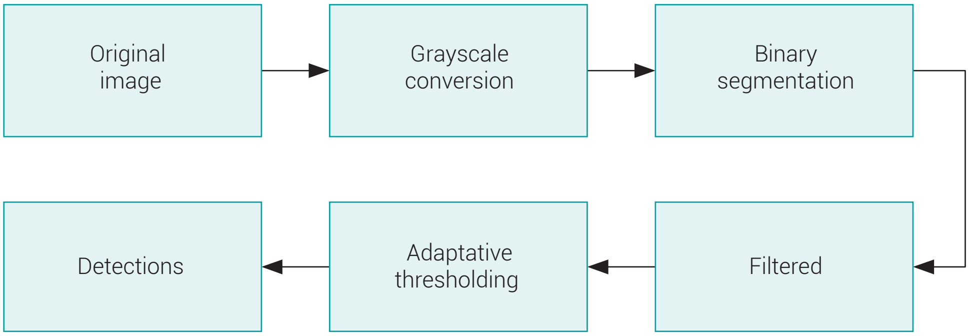

Figure 4. Block diagram of the background subtraction process.
Source: own work

Initially, the image obtained by the video camera is converted to grayscale, and then it is segmented in a binary way in order to separate in advance the image background and the main frame. Then, Gaussian blur filtering and morphological operations are performed to facilitate the adaptive process of thresholding with search of external contours. In this way, the objects in movement, present in the main frame of the image, are differentiated.

The grayscale conversion is obtained by estimating values between 0 and 255 to the pixels by behavior patterns[30][31], as shown in Equation 1.

$$y = y = (b * 0.11) + (g * 0.59) + (r * 0.3) \tag{1}$$

Subsequently, the creation of a mask based on regions of interest with an initial threshold value of 0.3 is proposed as a basic segmentation method for obtaining the binary image [32].

Likewise, the image obtained is filtered by two methodologies: Softening or Gaussian blurring [33], and by morphological operations [34]. Equation 2 shows the way in which Gaussian blurring is performed, where the values of x should fluctuate between -3 and 3 times the value of the standard deviation of the pixels, thus avoiding the tendency to infinity towards the negative side of the plane [35]. Equation 3 shows the mathematical expression related to the morphological operation of closure, as an improvement to the basic operations of dilation and erosion, so that an image A dilates and later erodes with respect to a structural element B.

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{\frac{x^2}{2\sigma^2}} \tag{2}$$

$$A \bullet B = (A \oplus B) \ominus B \tag{3}$$

Once the image has been filtered, a second threshold process is carried out, but this time by the adaptive method of Otsu to distinguish the moving objects that are candidates for people according to the size they occupy in the image [36]. Equation 4 shows how the optimal threshold value for distinguishing people is obtained, where L refers to the number of brightness levels.

$$\sigma_B^2(t^*) = max_{0\leq t\leq L-1}\,[log_2 \sigma_B^2(t^*)] \tag{4}$$

To denote the limits of the detected objects, the simple method of approaching and searching for contours is used, which manages to reduce the number of contours from 734 to 4, decreasing the memory space required during processing [37].

3.2.2 YOLO algorithm

YOLO is based on the use of convolutional neural networks to detect objects in images and videos by deep learning [38]. The YOLOv3 version is used for its fast object detection and real-time processing capability. Figure 3 shows the neural network architecture on which YOLO is based [39].

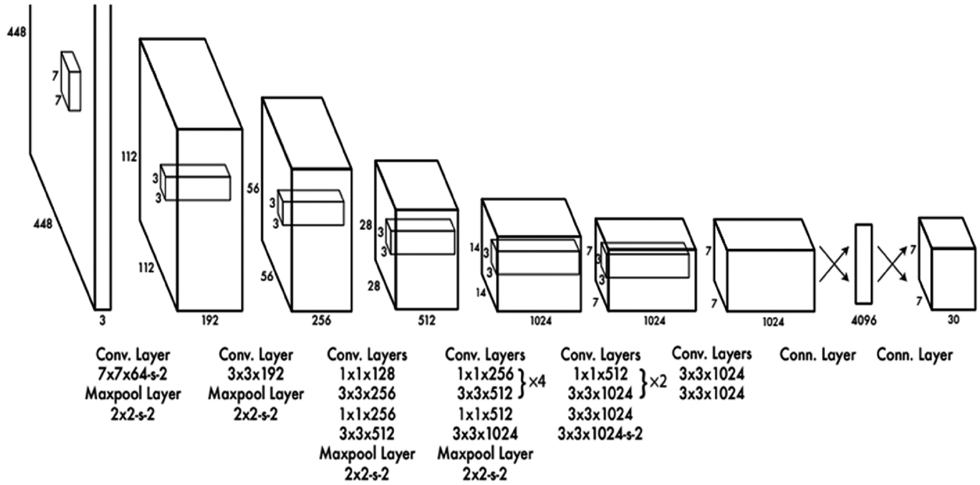

Figure 5. Architecture of the YOLO algorithm.
Source: adapted from [39]

The YOLO algorithm is based on the GoogLeNet classification model [40], and has 24 convolution layers and 2 connection layers [9]. For the present case, the algorithm is trained according to the COCO dataset, which has 80 object categories and more than 330,000 images of which approximately 200,000 are tagged. Because of the number of images provided by the data set and the fact that one of the 80 categories are people, it is not necessary to create a new data set or apply data enhancement techniques [41].

The methodology proposed for the implementation of the YOLO algorithm consists of 4 stages. In the first stage, the initial configuration of the files required for the implementation of YOLO is carried out. In the second stage, the image is converted to grayscale and then decomposed into the R, G and B channels, in order to know characteristic information according to the color scales [42]. In the third stage, the network input is configured by transforming the image into a blob. In the fourth stage, the Non-Maximum Suppression (NMS) algorithm is applied, so that redundancies in detection are eliminated [43]. The methodology is shown in Figure 6.

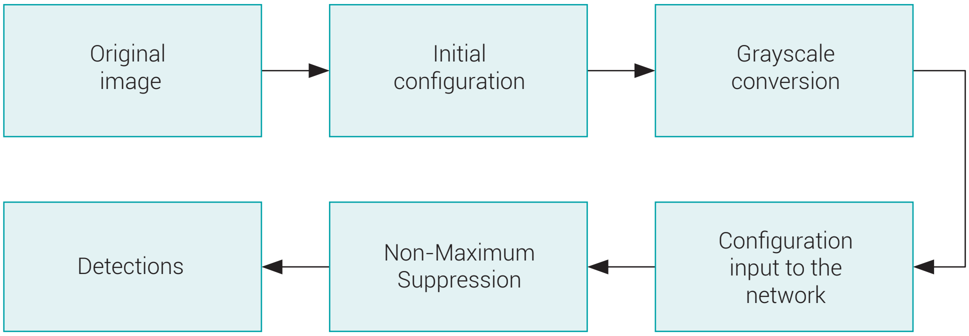

Figure 6. Block diagram of the YOLO algorithm process.
Reference: the authors

The files required for the implementation of the YOLO algorithm are those of weights, categories and configuration [44], which are available on the GitHub collaborative development platform. [45]. First, an initial configuration is required for the system with proposed values, which are shown in Table 1. The values referring to the number of divisions, subdivisions, size and other configuration values of the captured image are defined. Likewise, values are defined referring to learning rates of the neural network.

Table 1. Initial configuration of the parameters required by the YOLO algorithm.

Initial configuration of the YOLO algorithm for people detection			
Parameter	**Value**	**Parameter**	**Value**
Batch	64	Saturation	1.5
Subdivisions	16	Exposure	1.5
Width	512	Hue	0.1
Height	512	Learning rate	0.0001
Channels	3	Burn in	1,000
Momentum	0.9	Max batches	(500,200)
Decay	0.00005	Policy	(400000, 450000)
Angle	0	Scales	(0.1, 0,1)

Source: own work

Like the background subtraction technique, the YOLO algorithm requires a pre-processing stage where the image is converted to grayscale. In addition to this, the processing proposes the separation of the red, green and blue channels of the image, in order to know characteristic information of the image such as trends in the tone, [46], relevant when defining neural network input parameters.

Neural network input data are obtained from the conversion of image information into a blob [47]. Table 2 shows the configuration of the neural network input data for the third stage of the detection process. The required values are those referring to normalization scale, image size, average for the subtraction of pixels to the three image channels, and configuration of the order of the channels.

Table 2. Neural network input configuration

Configuration input to the network				
Scale	**Size**	**Mean**	**SwapRB**	**Crop**
0.00392	(416,416)	(104,117,123)	True	False

Source: own work

Once the input of the neural network is configured and the processing is executed, a confidence value is generated in the detection of the candidate objects [48]. The confidence value used is 0.6 and functions as a threshold, so that detections above this value are categorized as persons, and those below the threshold are omitted from the output window.

In learning detection processes, it is common to find redundancies in the detections, thus increasing the number of false positives in the process. For this reason, the Non-Maximum Suppression (NMS) algorithm is used as a post-processing technique [49][50], thus reducing exit assumptions and, therefore, redundancies in detection. For this process, an NMS threshold value of 0.5 is used.

3.3 Comparison of techniques

The comparison between the techniques for people detection in uncontrolled environments is made as a function of the response time of each algorithm and its effectiveness with respect to the number of hits in the detections. For the measurement of the response time, the task manager tool available on the Raspberry Pi 3B+ board is used, while the hit rate of the detections is obtained according to the mean square error with respect to the number of false positives and false negatives.

4. RESULTS

The video capture device, a 5MP Raspberry Pi camera, is located at a height of 4.2 meters with a line of sight to the circulation of people, in the Aula Sur F building of the Francisco de Paula Santander University, as shown in Figure 7.

Figure 7. Captured images.
Source: own work

Figure 8 illustrates the images resulting from the background subtraction process. First, the grayscale conversion is shown. Then, the binary segmentation of the image is presented, as well as the result of the filtering both by Gaussian blurring and by morphological closing operation. Finally, the detections obtained when performing an adaptive thresholding process by the Otsu method with a simple contour approximation are shown. Only the moving objects that circulate in the main frame of the video image are displayed.

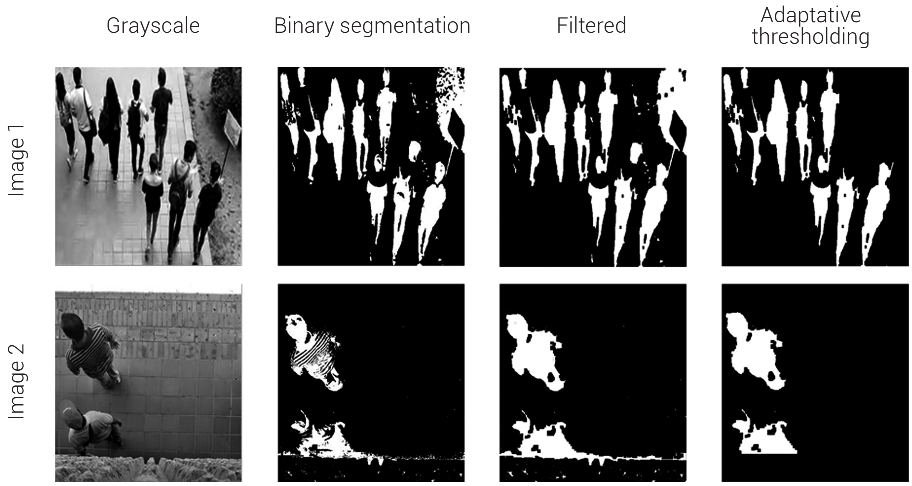

Figure 8. Background subtraction process.
Source: own work

The images obtained during the processing of the video image for the detection of people using the YOLO algorithm are also presented. Figure 9 shows the separation of the R, G and B channels after the conversion to grayscale.

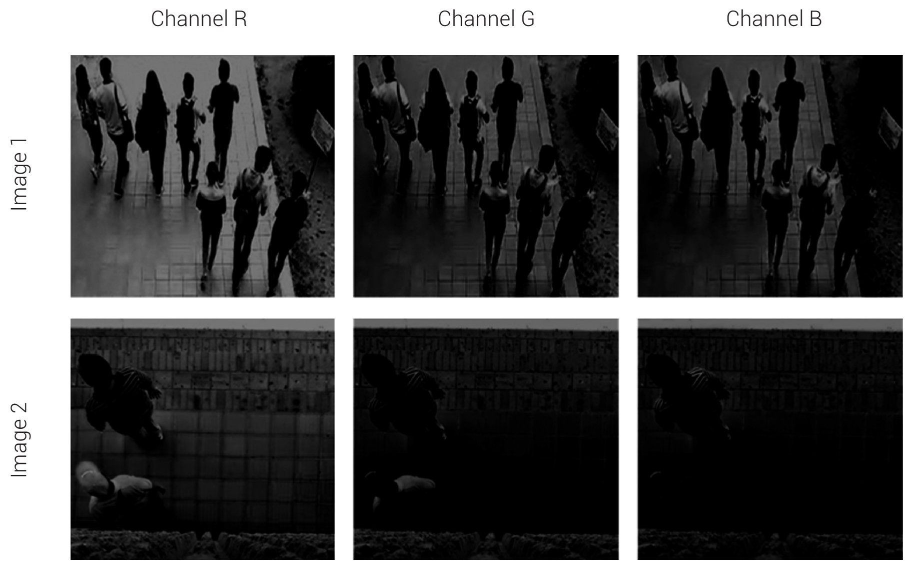

Figure 9. Separation of color channels in the image. From left to right the red channel, green channel and the blue channel.
Source: own work

From Figure 9 it is possible to infer a tendency to red in the captured images. In addition, figure 10 shows the images obtained during processing for the detection of persons using the YOLO algorithm. Redundancies in the detections are observed, so the need for a post-processing stage is inferred.

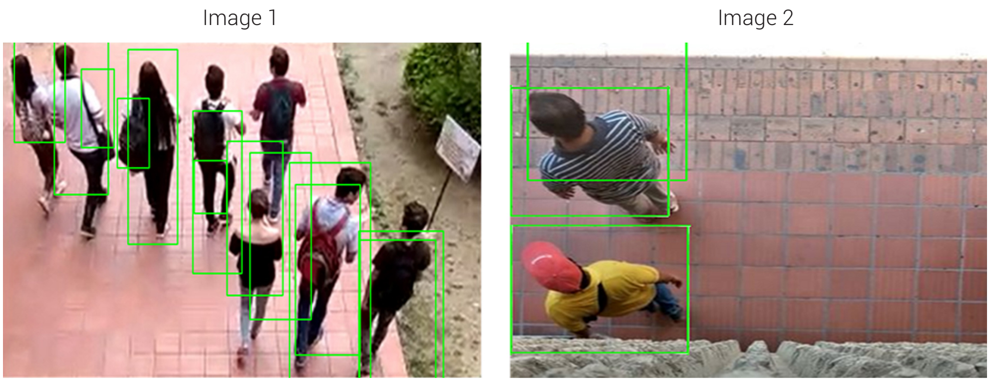

Figure 10. YOLO algorithm detections.
Source: own work

Detection redundancies are eliminated by applying the Non-Maximum Suppression algorithm as a post-processing method. Figure 11 shows cleaner detections, without redundancy, which improve system performance as the false positive rate of processing is reduced.

Image 1 Image 2

Figure 11. Application of the Non-Maximum Suppression algorithm to the detections obtained through the YOLO algorithm.
Source: own work

Once the application of the background subtraction technique and the YOLO algorithm for the detection of persons is completed, the comparison between the techniques is made with respect to the response time and the rate of success in the processing. Figure 12 shows the comparison between the techniques according to the response time in the Pi 3B+ raspberry plate.

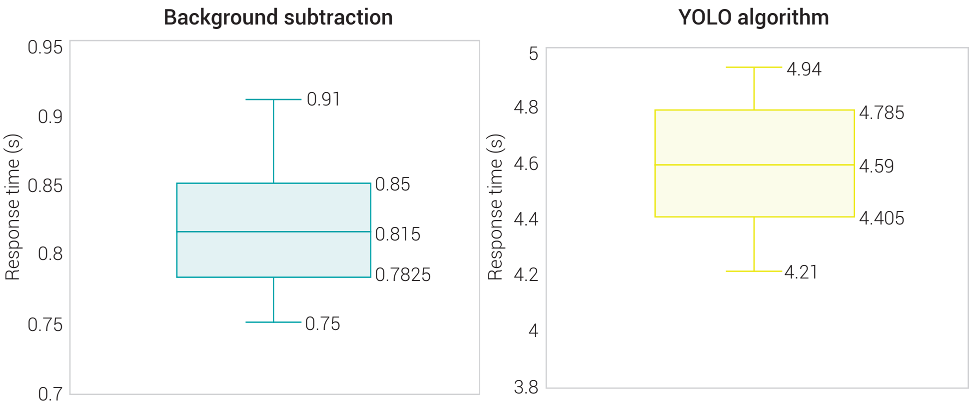

Figure 12. Response time comparison between background subtraction technique and YOLO algorithm.
Source: own work

For the background subtraction technique, the response time is between 0.75 and 0.91 seconds and has an average value of 0.815 seconds. Likewise, when using the YOLO algorithm, the average response time is 4.59 seconds, with values ranging from 4.21 to 4.94 seconds.

In addition, Table 3 presents the comparison between detection techniques according to the percentage of hits, with respect to the actual number of people circulating in the main frame of the video image. The values shown in the YOLO algorithm section refer to those obtained at the end of the post-processing stage through the Non-Maximum Suppression algorithm where redundancies are eliminated. The subtraction technique is a detection method based on the size of the objects, so it is not appropriate to apply a post-processing stage, as it is done in the YOLO algorithm, which performs detections regarding similarity of shapes.

Table 3. Successes of the people detection techniques in open spaces.

Technique	Real number of people	Number of detections	False positives	False negatives	Error
Background subtraction	32	28	1	5	15.93 %
YOLO algorithm	32	29	0	3	10 %

Reference: the authors

The background subtraction technique presented a mean square error of 15.93%, while this value when using the YOLO algorithm was 10%. These errors are mostly due to the number of false negatives when processing the captured images.

5. DISCUSSION

The detection of people in uncontrolled environments represents a challenge in people detection systems. The method proposed when using the background subtraction technique showed an error rate close to 16 %, due to factors such as dependence on the shade of the garments, for according to the method of segmentation of the background of the image used, the sequence is lost in the outline of the images and the same object is counted as multiple persons; or it is not catalogued as such. In addition, another factor that influences the efficiency of the detection system by background subtraction is the variation in the luminosity levels while the image is being captured. Therefore, an adaptive threshold stage was proposed by the Otsu method, thus allowing the coupling and differentiation of moving objects by mitigating the errors due to shadows generated by light. The proposed method can reduce the error rate by up to approximately 5% and is close to the values found in the literature in previous research if an adaptive method of segmentation of the image background has been added, as this would generate a greater hardware requirements; a relevant aspect if we take into account that the resources available on a Raspberry Pi 3B+ board are limited. Also, the YOLO algorithm presented a 10% error rate that is within the values handled in previous investigations. The main disadvantage of the proposed method with respect to the background subtraction technique is in the response time of the algorithm for real-time detection processes. It should be noted that the time obtained includes the time used in the post-processing stage by the Non-Maximum Suppression technique, as, although it is not within the basic stages of a detection process by the YOLO algorithm, it was additionally proposed in this research in order to mitigate errors due to overlapping in the image.

Similarly, the processes immersed in the proposed methodologies of the techniques in question can be improved if they are implemented in programming languages such as M (Matlab), focused on matrix manipulation, algorithms and information representation, as these also have specialized tools for signal and image processing. This would generate better performance in the detection process, but a greater investment for licensing costs in the implementation of them.

6. CONCLUSIONS

In the context of uncontrolled environments, detection processes become complex, causing hit rates to decrease regardless of the technique used. With the background subtraction technique, a hit rate of 84.07 % is obtained, while when using the YOLO algorithm, this value rises to 90 %.

The difference of approximately 6 % in the hit rate between the techniques used is due to factors such as overlapping, light changes and color dependence of the garments when using the background subtraction technique. These problems are mitigated with the YOLO algorithm.

The response time, when applying the background subtraction technique, was on average 5 times less than when applying the YOLO algorithm. From this, it is possible to recommend the use of the background subtraction technique in hardware tools such as the Raspberry Pi 3B+ board for processes in which the analysis of information in real time is prioritized, while the YOLO algorithm presents the characteristics required in processes where the information is analyzed after the acquisition of the image.

7. REFERENCES

[1] C. V. Niño Rondón, S. A. Castro Casadiego, B. Medina Delgado, and D. Guevara Ibarra, "Análisis de viabilidad y diseño de un sistema electrónico para el seguimiento de la dinámica poblacional en la ciudad de Cúcuta," *Ing. USBMed*, vol. 11, no. 1, pp. 56–64, 2020, doi: 10.21500/20275846.4489

[2] E. N. Kajabad and S. V. Ivanov, "People Detection and Finding Attractive Areas by the use of Movement Detection Analysis and Deep Learning Approach," *Procedia Comput. Sci.*, vol. 156, pp. 327–337, 2019, doi: 10.1016/j.procs.2019.08.209

[3] M. Hussein, W. Abd-Almageed, Y. Ran, and L. Davis, "Real-time human detection, tracking, and verification in uncontrolled camera motion environments," *Proc. Fourth IEEE Int. Conf. Comput. Vis. Syst. ICVS'06*, vol. 2006, no. February, pp. 1-7, 2006, doi: 10.1109/ICVS.2006.52

[4] A. K. Agrawal and Y. N. Singh, "An efficient approach for face recognition in uncontrolled environment," *Multimed. Tools Appl.*, vol. 76, no. 3, pp. 3751–3760, 2017, doi: 10.1007/s11042-016-3976-z

[5] B. Garcia-Garcia, T. Bouwmans, and A. J. Rosales Silva, "Background subtraction in real applications: Challenges, current models and future directions," *Comput. Sci. Rev.*, vol. 35, pp. 1-42, 2020, doi: 10.1016/j.cosrev.2019.100204

[6] E. S. Jeon *et al.*, "Human detection based on the generation of a background image by using a far-infrared light camera," *Sensors (Switzerland)*, vol. 15, no. 3, pp. 6763–6788, 2015, doi: 10.3390/s150306763

[7] Y. Yang, Q. Zhang, P. Wang, X. Hu, and N. Wu, "Moving Object Detection for Dynamic Background Scenes Based on Spatiotemporal Model," *Adv. Multimed.*, vol. 2017, pp. 1-14, 2017, doi: 10.1155/2017/5179013

[8] G. Natanael, C. Zet, and C. Fosalau, "Estimating the distance to an object based on image processing," *EPE 2018 - Proc. 2018 10th Int. Conf. Expo. Electr. Power Eng.*, pp. 211–216, 2018, doi: 10.1109/ICEPE.2018.8559642

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91

[10] J. Zuo, Z. Jia, J. Yang, and N. Kasabov, “Moving Target Detection Based on Improved Gaussian Mixture Background Subtraction in Video Images,” *IEEE Access*, vol. 7, pp. 152612–152623, 2019, doi: 10.1109/ACCESS.2019.2946230

[11] A. Miranto, S. R. Sulistiyanti, and F. X. Arinto Setyawan, “Adaptive background subtraction for monitoring system,” *2019 Int. Conf. Inf. Commun. Technol. ICOIACT 2019*, pp. 153–156, 2019, doi: 10.1109/ICOIACT46704.2019.8938501

[12] W. Liu, H. Yu, H. Yuan, H. Zhao, and X. Xu, “Effective background modelling and subtraction approach for moving object detection,” *IET Comput. Vis.*, vol. 9, no. 1, pp. 13–24, 2015, doi: 10.1049/iet-cvi.2013.0242

[13] S. N. N. Htun, T. T. Zin, and P. Tin, “Image Processing Technique and Hidden Markov Model for an Elderly Care Monitoring System,” *J. Imaging*, vol. 6, no. 6, pp. 1-23, 2020, doi: 10.3390/jimaging6060049

[14] M. Mariappan, L. K. Thong, and K. Muthukaruppan, “A design methodology of an embedded motion-detecting video surveillance system,” *Int. J. Integr. Eng.*, vol. 12, no. 2, pp. 55–69, 2020, doi: 10.30880/ijie.2020.12.02.007

[15] S. Li, M. O. Tezcan, P. Ishwar, and J. Konrad, “Supervised people counting using an overhead fisheye camera,” in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2019*, 2019, pp. 1–8, doi: 10.1109/AVSS.2019.8909877

[16] K. Rantelobo, M. A. Indraswara, N. P. Sastra, D. M. Wiharta, H. F. J. Lami, and H. Z. Kotta, “Monitoring Systems for Counting People using Raspberry Pi 3,” in *2018 International Conference on Smart Green Technology in Electrical and Information Systems: Smart Green Technology for Sustainable Living, ICSGTEIS 2018 - Proceeding*, 2018, no. October, pp. 57–60, doi: 10.1109/ICSGTEIS.2018.8709141

[17] P. Ren, W. Fang, and S. Djahel, “A novel YOLO-Based real-Time people counting approach,” in *2017 International Smart Cities Conference, ISC2 2017*, 2017, vol. 243, pp. 3–4, doi: 10.1109/ISC2.2017.8090864

[18] S. Vaidya, N. Shah, N. Shah, and R. Shankarmani, “Real-Time Object Detection for Visually Challenged People,” in *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, 2020, no. Iciccs, pp. 311–316, doi: 10.1109/ICICCS48265.2020.9121085

[19] A. Lucian, A. Sandu, R. Orghidan, and D. Moldovan, "Human leg detection from depth sensing," in *2018 IEEE International Conference on Automation, Quality and Testing, Robotics, AQTR 2018 - THETA 21st Edition, Proceedings*, 2018, pp. 1–5, doi: 10.1109/AQTR.2018.8402735

[20] M. Ahmad, I. Ahmed, and A. Adnan, "Overhead View Person Detection Using YOLO," in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2019*, 2019, no. i, pp. 0627–0633, doi: 10.1109/UEMCON47517.2019.8992980

[21] Y. Yang, K. Han, S. HyungLee, and J. Lee, "Temporal difference based adaptive object Detection (ToDo) platform at Edge Computing System," in *17th Annual Consumer Communications & Networking Conference (CCNC)*, 2020, pp. 1–2, doi: 10.1109/JPROC.2019.2918951

[22] C. Liu, P. Liu, and C. Xiao, "Detecting stabbing by a deep learning method from surveillance videos," in *IST 2019 - IEEE International Conference on Imaging Systems and Techniques, Proceedings*, 2019, pp. 1–5, doi: 10.1109/IST48021.2019.9010206

[23] A. Aktas, B. Dogan, and Ö. Demir, "Derin öğrenme yöntemleri ile dokunsal parke yüzeyi tespiti," *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Derg.*, vol. 3, pp. 1685–1700, 2020, doi: 10.17341/gazimmfd.652101

[24] R. Bohush and I. Zakharava, "Person tracking algorithm based on convolutional neural network for indoor video surveillance," *Comput. Opt.*, vol. 44, no. 1, pp. 109–116, 2020, doi: 10.17341/gazimmfd.652101

[25] F. Fang, K. Qian, B. Zhou, and X. Ma, "Real-time RGB-D based people detection and tracking system for mobile robots," in *2017 IEEE International Conference on Mechatronics and Automation, ICMA 2017*, 2017, pp. 1937–1941, doi: 10.1109/ICMA.2017.8016114

[26] G. Cavanzo, M. Pérez, and F. Villavisan, "Medición de eficiencia de algoritmos de visión artificial implementados en raspberry pi y ordenador personal mediante Python," *Ingenium*, vol. 18, no. 35, pp. 105–119, 2017, doi: 10.21500/01247492.3218

[27] S. Sivaranjani and S. Sumathi, "A review on implementation of bimodal newborn authentication using raspberry Pi," in *Global Conference on Communication Technologies, GCCT 2015*, 2015, no. Gcct, pp. 267–272, doi: 10.1109/GCCT.2015.7342664

[28] R. Barbuzza, J. D'amato, A. Rubiales, L. Dominguez, A. Perez, and M. Venere, "Un Método para la Sustracción de Fondo en Videos Inestables," *Mecánica Comput.*, vol. 34, no. 51, pp. 3409–3417, 2016, doi: 10.13140/2.1.1478.1123

[29] C. V. Niño Rondón, S. A. Castro Casadiego, and B. Medina Delgado, "Caracterización para la ubicación en la captura de video aplicado a técnicas de visión artificial en la detección de personas," *Rev. Colomb. Tecnol. Av. RCTA*, vol. 2, no. 36, pp. 83–88, 2020, doi: 10.24054/16927257.v36.n36.2020.3720

[30] G. Sánchez-Torres and J. A. Taborda-Giraldo, "Estimación automática de la medida de ocupación de playas mediante procesamiento de imágenes digitales," *TecnoLógicas*, vol. 17, no. 33, pp. 21-29, 2014, doi: 10.22430/22565337.543

[31] D. Jyoti Bora, R. Kumar Bania, and C. Ngoc, "A Local Type-2 Fuzzy Set Based Technique For He Stain Image Enhancement," *Ing. Solidar.*, vol. 15, no. 29, pp. 1–22, 2019, doi: 10.16925/2357-6014.2019.03.02

[32] C. T. Hsieh, H. C. Wang, Y. K. Wu, L. C. Chang, and T. K. Kuo, "A Kinect-based people-flow counting system," in *ISPACS 2012 - IEEE International Symposium on Intelligent Signal Processing and Communications Systems*, 2012, no. Ispacs, pp. 146–150, doi: 10.1109/ISPACS.2012.6473470

[33] C. S. Marzan and N. Marcos, "Towards tobacco leaf detection using Haar cascade classifier and image processing techniques," *ACM Int. Conf. Proceeding Ser.*, pp. 63–68, 2018, doi: 10.1145/3282286.3282292

[34] B. Rodríguez-Cuenca, S. García-Cortés, C. Ordóñez, and M. C. Alonso, "Morphological operations to extract urban curbs in 3d mls point clouds," *ISPRS Int. J. Geo-Information*, vol. 5, no. 6, pp. 1-17, 2016, doi: 10.3390/ijgi5060093

[35] E. Elboher and M. Werman, "Efficient and accurate Gaussian image filtering using running sums," in *International Conference on Intelligent Systems Design and Applications, ISDA*, 2012, pp. 897–902, doi: 10.1109/ISDA.2012.6416657

[36] P. Huamaní Navarrete, "Umbralización múltiple utilizando el método de Otsu para reconocer la luz roja en semáforos," *Scientia*, vol. 17, no. 17, pp. 247–262, 2016, doi: 10.31381/scientia.v17i17.393

[37] G. Xu, Z. Wang, Y. Cheng, Y. Tian, and C. Zhang, "A man-made object detection algorithm based on contour complexity evaluation," *Chinese J. Aeronaut.*, vol. 30, no. 6, pp. 1931–1940, 2017, doi: 10.1016/j.cja.2017.09.001

[38] X. Zou, "A Review of object detection techniques," in *Proceedings - 2019 International Conference on Smart Grid and Electrical Automation, ICSGEA 2019*, 2019, pp. 251–254, doi: 10.1109/ICSGEA.2019.00065

[39] M. Massiris, C. Delrieux, and J. Á. Fernández, "Detección de equipos de protección personal mediante red neuronal convolucional YOLO," no. September, pp. 1022–1029, 2020, doi: 10.17979/spudc.9788497497565.1022

[40] C. Szegedy *et al.*, "Going deeper with convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1–9, 2015, doi: 10.1109/CVPR.2015.7298594

[41] J. Hu, X. Gao, H. Wu, and S. Gao, "Detection of Workers Without the Helments in Videos Based on YOLO V3," in *12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2019, pp. 1553–1560, doi: 10.1109/CISP-BMEI48845.2019.8966045

[42] J. Yang, Q. Min, W. Lu, Y. Ma, W. Yao, and T. Lu, "An RGB channel operation for removal of the difference of atmospheric scattering and its application on total sky cloud detection," *Atmos. Meas. Tech.*, vol. 10, no. 3, pp. 1191–1201, 2017, doi: 10.5194/amt-10-1191-2017

[43] R. Rasmus, G. Matthieu, and V. G. Luc, "Non-Maximum Suppression for Object Detection by Passing Messages between Windows," in *Asian Conference on Computer Vision*, 2014, vol. 9003, pp. 1-16, 2014, doi: 10.1007/978-3-319-16865-4

[44] R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," in *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 2019, pp. 2503–2510, doi: 10.1109/BigData.2018.8621865

[45] P. Tumas, A. Nowosielski, and A. Serackis, "Pedestrian Detection in Severe Weather Conditions," *IEEE Access*, vol. 8, pp. 62775–62784, 2020, doi: 10.1109/ACCESS.2020.2982539

[46] R. Gothwal, S. Gupta, D. Gupta, and A. K. Dahiya, "Color image segmentation algorithm based on RGB channels," in *Proceedings - 2014 3rd International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions, ICRITO 2014*, 2015, pp. 1–5, doi: 10.1109/ICRITO.2014.7014669

[47] Y. Xu, T. Wu, F. Gao, J. R. Charlton, and K. M. Bennett, "Improved small blob detection in 3D images using jointly constrained deep learning and Hessian analysis," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1109/ICRITO.2014.7014669

[48] S. Shinde, A. Kothari, and V. Gupta, "YOLO based Human Action Recognition and Localization," *Procedia Comput. Sci.*, vol. 133, no. 2018, pp. 831–838, 2018, doi: 10.1016/j.procs.2018.07.112

[49] R. Seidel, A. Apitzsch, and G. Hirtz, “Improved Person Detection on Omnidirectional Images with Non-maxima Supression,” *VISIGRAPP 2019 - Proc. 14th Int. Jt. Conf. Comput. Vision, Imaging Comput. Graph. Theory Appl.*, vol. 5, no. January, pp. 474–481, 2019, doi: 10.5220/0007388404740481

[50] D. Ramírez-Jiménez and J. D. Quintero-Ospina, “Clasificación de patologías presentes en la columna vertebral mediante técnicas de máquinas de aprendizaje,” *Ing. Solidar.*, vol. 15, no. 27, pp. 1–23, 2019, doi:10.16925/2357-6014.2019.01.05