SYSTEM FOR PEOPLE COUNTING BY COMPUTER VISION

Sistema para conteo de personas basado en visión por computador

Jhonny Santiago Chamorro Rodriguez
Universidad Técnica del Norte
GISI
Ibarra, Ecuador
jschamorro@utn.edu.ec

***Abstract*—The detection and counting of people is a method widely used for monitoring and video surveillance. The objective of this project is to perform the counting of people process from flat images, with a camera located on top of an interior environment. The algorithm used for the counting of people process was developed by morphological processes based on flat images, since the camera is on top of an interior environment, the same ones that serve to modify the system behaviour. The counting system is divided into two components; one component is for image monitoring and the other one for image acquisition.**

The system used for counting people has some conditions that need to be met in order for it to work properly; First of all, it is important to check if the person is in the detection zone; secondly, the system has limits at the top and bottom areas to check if the person enters or leaves the detection zone. The people counting system was coded in Python through the OpenCv library which shows the data coming from the camera, indicating the number of people who entered or left a particular site. The algorithm that was developed for this counting system was evaluated in a classified and registered database.

I. Introduction

Currently the applications that are based on the counting of people are very used for which there is a great demand, due to its system of security and video surveillance that allows to know the amount of people who enter and leave of place cite Head.

Some methods for detecting people are based on background subtraction, in the tracking of objects in the foreground is done through the occupation area to differentiate the number of people and a color vector to track each person. Another type of system is where the camera is inclined, these systems need to obtain a perspective of the person to carry out their detection cite doc: Motion.

Most systems that perform the counting of people do so with a single camera that is oriented in a zenith and at a certain height. The system has two regions of interest defined at the top and bottom of the image. The histogram of these regions of the motion picture is calculated and by comparing a threshold detect the occupancy level. The bi-directional count is based on a deterministic algorithm that is based on the crossing of the upper and lower regions. This type of system with a zenithic orientation can not perform an analysis of the detected object, since the view does not offer enough characteristics to detect the person.

II. Methodology

A. Segmentation

B. Background extraction

The MOG background extractor was used based on the concordance of each of the Gaussians of the mixture, it is determined if they correspond to the background colors. Then the values of a particular pixel are modeled as a mixture of Gaussian distributions and pixel values that do not correspond to background colors are considered foreground until there is a Gaussian that includes them.

First, it must be known that if a pixel results from a particular surface under a given illumination, a simple Gaussian distribution would be sufficient to model the value of the pixel. On the other hand, if only changes of illumination take place during hours, a simple and adaptive Gaussian distribution would suffice. But in practice, multiple surfaces often appear in the view of a particular pixel and lighting conditions change. Thus, in this case, adaptive multiple gaussians are needed. This method then uses an adaptive Gaussian mixture to approximate this process. The following are conclusions of the algorithm used:

- Background of the scene (any object that after a time is included in the background): The algorithm has optimal response with the parameters set in indoor scenes.
- Dynamic background: On many occasions, the dynamic background is detected as a foreground.
- Lighting changes: In this case occasionally if the areas affected by the lighting changes are detected as foreground.
- Static Foreground: In the case of when the person is stopped and makes small movements in the detection area the algorithm has a shorter response time and the behavior is better.

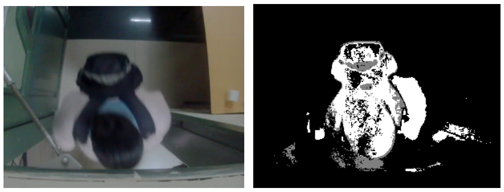

Fig. 1. Response of the MOG algorithm to dynamic foreground

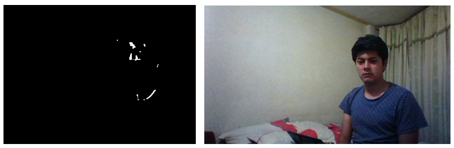

Fig. 2. MOG algorithm response to static foreground

In this case tests were performed at medium and short distances, at these distances the algorithm does not have problems for the detection of movement, in half distances the algorithm presents greater noise due to the shadow that is generated and therefore will detect two people either one on top of the other or someone with higher height, but its correct operation would be at short distances because it presents less noise and better behavior at the moment of detecting motion [12].

III. Morphological filter

As explained in the previous chapter on the morphological filter and the operations that can be applied to solve the problem we have that morphological transformations aim to modify the shape of objects that are observed in an image. In this application it is very necessary, since the foreground image resulting from the previous module contains parts that are affected by the generation of regions or pixels with noise. Therefore, you can have fake motion zones where there is really no movement, or when the movement is minimal, so you would be considered a negligible region. To solve this problem, the erosion and dilation operations of morphological filtration are applied.

A. Erosión

It is the progressive degradation of one of the fields that can be 0 or 1. An element of the field to be degraded will continue to belong to it if it is surrounded by similar elements, otherwise it will pass to the other field. Mathematically it will be expressed as follows: if we take the symmetric structural element with respect to the origin of B, the erosion of a set X with respect to the element B is:

$$X \ominus B = \{B_X \subset X\} \tag{1}$$

What is equal to a successful transformation or fails where B is the empty set. The symbol $\ominus$ represents the subtraction of Minkowski.

B. Dilatation

It is the progressive growth of one of the fields (0 or 1). An element of the opposite field to grow will be converted if it has

Fig. 3. Example aplying erosion

a neighbor belonging to the field that is expanded. Otherwise, it will remain the same. The elements belonging to the field to be expanded are evidently unchanged. If you applied a large number of times would end up destroying the image since all the pixels would be high. Mathematically, dilation can be expressed as:

$$X \oplus B(X^C \ominus B^C) \tag{2}$$

Fig. 4. Example applying expansion

C. Combination of both

One of the main characteristics of the expansion and erosion is that they do not comply with the commutative property, that is, it is not the same to realize a dilation followed by an erosion, than an erosion followed by a dilation. In this project, two replications are performed for erosion first and then one for dilatation. When the operations are performed in this order the process receives the name of opening. It is denominated well since to begin with an erosion tends to break the pieces in its constituent parts. The main reasons why this order is chosen and not the reverse is that it softens the contours of the object, breaks thin bonds and removes small protuberances.

IV. Contour detection

In each frame are grouped moving pixels that are continuous with each other as a single region. Next we have that the image in which the white pixels are being worked represent the zones that are in movement and the blacks represent the immobile zones. Since the foreground mask is a binary image it is not necessary to convert the image that has several gray levels into a new one with only two (Umbralizacin). There is a method in which it follows the image starting from the upper left corner to the right until the image is finished, when the first pixel that is at the high level is located it is assigned a 1, indicating that it was found the first object. We continue to check carefully if the neighbors are also in high level, if they are they are assigned the same number. Another case is when a pixel is not neighbor of one already previously labeled and if it is in high level, it is assigned the next number that would be 2 respectively.

Fig. 5. Contour of a person

V. Detection of people

From the contour detection it is determined whether it belongs to people or not, taking into account that it requires detection of target people. First filtering of areas is done, taking into account that areas that are above or below a value are filtered. So much that the main objective is the detection and counting of people in a certain range of distance, and possesses certain limits of major or minor areas. Then the next filter is applied that has to do with the long and wide relation of the person detected, whether it is standing or sitting, if in a case it is required to enter people in a row holding hands would expand the limit already be higher or lower to avoid detection conflicts. Due to the restriction of limits that it possesses it can be limited the detection of objects that have characteristics similar to the person in relation to its width and the proportion of its stop.

It is also necessary to take into account in the area of detection that it encompasses the contour of the person and if it does not it would not detect the person or would cause false detections. Another important point is the resolution of the video, since the lower resolution of the video has a better detection and counting compared to high resolutions. With everything mentioned above you get a very good percentage of correct detections, but also you get some margin of error therefore it is not totally reliable.

Finally, after explaining how people were detected, it should be emphasized that the program is designated to the count of a single person and the process consists of detecting the person's position and locating each frame of the person. On the other hand, when there is more than one person in the detection area, the program can not detect x people in x positions [13].

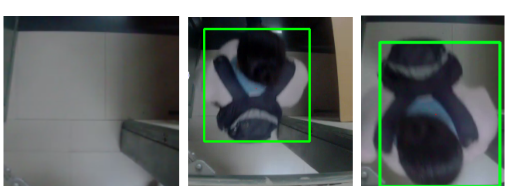

Fig. 6. Detection of a person

VI. Video output

This section is the one that deals with the generation of the output video, being the final result of the application. The output video is constantly executed and it shows each result at the moment of its operation. This process starts in the obtaining of images by means of a camcorder, that step is simple since it is realized automatically. The application consists of two windows, the first window shows the input image either in the form of recorded video or in real time, taking into account that the detection area encompasses the person detected in each position. When having the window in which the entry and exit of people is visualized, the moment in which the person is detected is observed and when the limits are passed, the count is higher or lower, this is done to improve the visualization of the user. It is done in order for the user to visualize their area of video surveillance and is aware of what happens at that time and how the algorithm behaves in environments and internal scenes. The number of people detected at each instant, both incoming and outgoing, is displayed in the same way, the information is stored and then used as desired by the user. Finally, it can also be observed in the second window when the morphological filter is applied, with this the person can have evidence of what happens in the application and if the detection and counting of people is accurate or not.

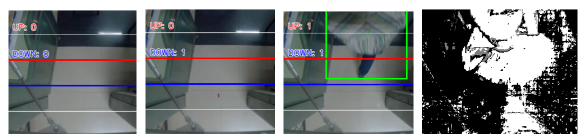

Fig. 7. Detection of a person

Tests and results

For the tests and results were recorded sequences of images with different resolutions with a camera Motion Pro Action Camera which allows us to record in different resolutions the video, the camera is located at a height of (2.30m) and 90° with soil reference, in an indoor environment. In the first video was recorded at a resolution of 480x480 pixels, and twenty four frames per second. The implementation of the person counter performed on Fig VI resulted in the following:

Fig. 8. Resolution video images 480x480 pixels

- A video was recorded in which the passage of people is observed both entering and leaving, in that video enter 4 people and leave 4, in which you get false positives, false negatives and correct detection of people. To improve the count, the area of interest is enlarged thus reducing false positives and negatives in an interior environment.

- In this image representation, three persons were detected for the entry, in addition to detecting four false positives and a false negative representing the error in the count of a person.
- For the output, four people were detected, four false positives and no false negatives. It has that in the exit there is less error as far as the counting of people but not in relation to the false positives and negatives.

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	5	5	0	10
Ent.6	4	5	1	10

TABLE I
FIRST VIDEO RESULTS

In the second video was recorded at a resolution of 320x240 (VGA) pixels and thirty frames per second. The implementation of the people counter performed on fig VI resulted in the following:

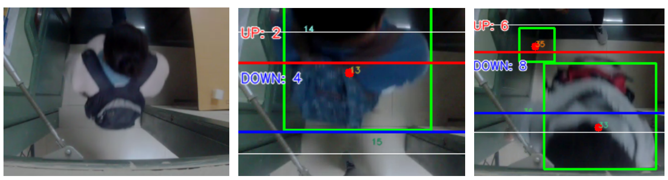

Fig. 9. 320x240 pixel resolution video images(VGA)

- A video was recorded in which you can see the passage of men and women entering and leaving the environment, the video has to enter 6 people and come out 5 in which you get false positives, false negatives and the correct detections . For the improvement of the count, the area of interest for the count is enlarged thus reducing false positives and negatives in the proposed scene.
- In this representation of images you have that for the entrance five people were detected in addition to detecting eight false positives and a false negative that represents the error in the count of a person.
- For the output it was obtained that five people were detected, three false positives and no false negative. It has that in the exit there is less error in the counting of people referring to the characteristics of the video.

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	5	3	0	8
Ent.6	5	8	1	13

TABLE II
VIDEO SECOND RESULTS

The third video was recorded at a resolution of 320x240 pixels (VGA) and thirty frames per second. The implementation of the people counter performed on fig VI resulted in the following:

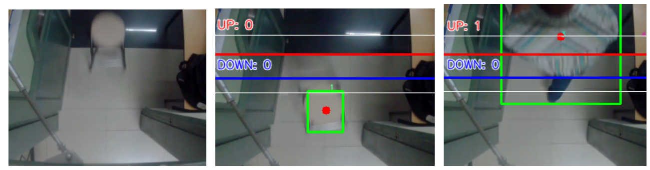

Fig. 10. 320x240 pixel resolution video images(VGA)

- In this video were realized tests related to counting of objects and people, the result that was obtained was that the counting of people and objects, but to correct said error what was done was to reduce the ranges between line of entry and line of exit , in doing this you have that the object is not counted but if the person.
- In this representation of images one has to enter six people and five people leave, five people were detected for the entrance but three false positives and a false negative were also detected.
- For the output it was obtained that five persons, seven false positives and no false negatives were detected. It has that in the exit there is less error in the counting of people but you have more false positives than in the entrance.

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	5	7	0	12
Ent.6	5	3	1	9

TABLE III
THIRD VIDEO RESULTS

In this video the same resolution spoken in the first video is present but with different conditions, therefore you will get different results. The implementation of the people counter performed on fig VI resulted in the following:

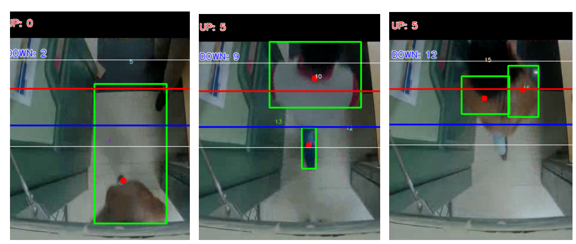

Fig. 11. 320x240 pixel resolution video images(VGA)

- In this video tests were performed referring to counting of objects and people in which the ranges of area of interest are the same tested previously but with different results.
- In this representation of images you have to enter four people and leave four people, two people were detected for the entrance but also detected a false positive and two false negatives.
- For the output it was obtained that three persons, eight false positives and one false negative were detected. You have that both in the entrance and in the exit you have more counting errors than the previous tests.

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	4	8	2	14
Ent.6	3	1	1	5

TABLE IV
THIRD VIDEO RESULTS

CONCLUSIONS

- The present work presents a counting system of people through a camera located in a zenith, differentiating between counting of entrance and exit. It proposes a model for the detection of heads, softening the image to eliminate noise and erroneous detections, and then use the Canny filter as a basis for counting and tracking people. By means of the Canny filter specifications it has good detection, good localization and minimum response, thus solving temporal occlusion problems that occur in test videos.
- The application presents an autonomous operation since the user at no time manipulates variables during the execution of the application. It also has a low cost in the acquisition stage, and it was demonstrated that using a low quality camera the system has a good operation, and the acquisition of images is done in real time and through previously recorded videos, it is better perform it in real time to avoid loss of information and false detections.
- The subtractive technique of the background allowed to obtain a better detection of movement since it allows to segment objects of interest that are in the scene. With a resolution of 160x120 pixels, this gives enough information to detect people passing through the counting area. In addition, a methodology was implemented to monitor the people recognized in the video sequence, by predicting the movement of their centroids so that passers do not have the same pattern between frame and frame, additionally monitoring and detection helps to have a record the amount of people entering or leaving the scene.
- The system presented some errors regarding image quality where there are false detections when there is a considerable difference in the contrast between the background and the person, the solution for this is to reduce the quality of the video until you have an adequate number of frames per second, thus avoiding errors and improving the operation of the system.

REFERENCES

[1] Z. Z. Y. X. Bin Li, Jian Zheng, “A People Counting Method Based on Head Detection and Tracking”, Jul. 2011.

[2] S. S. Sumati Manchanda, “Analysis of Computer Vision based Technique for Motion Detection”, Jun. 2016.

[3] R. G. Jiménez, “Detección y conteo de personas, a partir de mapas de profundidad cenitales capturados con cámaras TOF”, Dec 2015.

[4] E. M. J. Diego Alberto Vera Meaurio,“Conteo de personas en imagen y video mediante la técnica de Viola-Jones a travéz de clasificadores Haar utilizando software libre”, Feb. 2014.

[5] C. S. K. Fabiana da Silva Guizi, “Real-time People Detection and Tracking using 3D Depth Estimation”, Apr. 2016.

[6] K.-W. L. Young-Ho Suh, Sang Keun Rhee, “Continuous Location Tracking of People by Multiple Depth Cameras”, Sep. 2015.

[7] J. M. ValenciaLópez, “Registro de transeúntes en tiempo real utilizando un sistema de visión artificial sobre un ambiente controlado”, 2007.

[8] H. Ma and M. Z. Huchuan Lu, A Real-time Effective System for Tracking Passing People, Jul. 2008.

[9] I. A. S. M. Analí J.Alfaro Alfaro,“ Diseño de un Algoritmo de Segmentación de Imágenes aplicando el Funcional de Mumford-Shah para mejorar el desempe no de los Algoritmos Clásicos de Segmentación”, Mar. 2010.

[10] R. B. L.Gervasoni, J.P.D amato, “Un Método eficiente para la sustracción de fondo en videos usando Gpu”, Sep. 2014.

[11] R. G. y R.e.Wood, Digital Image Processing, Dec. 1993.

[12] D. Lefloch, “Real-Time People Counting system using Video Camera”. Jan, 2007.

[13] U. Gonzáles Benítez, “Detección de personas a partir de visión artificial”, Sep. 2010

[14] G. Córdova Manobanda, “Recuento de personas en espacioscerrados mediante sensores RGB+D”. *España*, 2016.

[15] [Online]. Available: http://docs.opencv.org/master/d9/df8/tutorial/root.html #Opencvmaster