

# Pedestrians detections by adaptative background mixtured model and histogram of oriented gradients

Otho Teixeira Komatsu  
Department of Computer Science  
University of Brasília  
Brasília, Brasil  
otho.tk@hotmail.com

Giordano Süffert Monteiro  
Department of Computer Science  
University of Brasília  
Brasília, Brasil  
giordano@suffert.com

**Abstract**—The need of a technology based on pedestrian detection and models to describe a scene from videos has been largely a research topic, bringing out a diversity of techniques and tools to improve the process. In this report, the algorithm was based on a improved adaptative background mixtured model, a technique that allows the program to detect distinguish between the moving objects and the background from the scene. To detection of human, histogram of oriented gragients was implemented along with the process of nonmax supression, classifying and tracking through the frame using a pre-trained Support Vector Machine.

**Index Terms**—Adaptative background mixtured model, Histogram of oriented gragients,Support Vector Machine .

## I. INTRODUCTION

Detection technology is a vast area that is studied and researched for several years, since its applicability is always needed on computer vision problems, e.g. surveillance cameras analysis and automated machines based on detection. As a result, a number of techniques and different approaches are available to solve human and pedestrian detection problem. Most of them share some common approach and ways of solutions.

Throughout the problem solving process and the search for a fine model to work around some problems that came up in the implementation, it was concluded that the use of Adaptative Background mixtured model, proposed by P. KaewTraKulPong *et al* [2], required to separate the foreground(moving objects from the scene) from the background(stationary objects), and Histogram of oriented gradients[3] (together with support vector machine[4]), on the other hand, to measure features from a window of the scene to analysis and detection of human, was a reasonable approach to the algorithm's computation time and performace comparing to other approaches, results concluded from the papers that this report is based on.

Some additional methods was implemented in the program with the objective of include some solutions to unfavorable environments of detection. The movement of background objects, such as tree's leaves, and equipment failures in the process of filming, as camera's automatic focusing, was some of reasons that was proposed some alternative solutions to these adverse events.

## II. BACKGROUND AND RELATED WORK

Xin Zhang *et al* proposed the process of Multi-Frame-differencing followed by an adaptative background model to produce the tracking of moving regions and objects from a scene.

P. KaewTraKulPong *et al* proposed an specific method of calculate the adaptative background model with an algorithm based on Gaussian distribution as mixtured model to estimate a model pixel to the update of each frame in the subtraction process, along with the online expectation maximisation algorithm. His algorithm also permits the shadow detection from scenes, reducing false-positives erros during the detection.

Navneet Dalal *et al* proposed an histogram of oriented gradients model to collect relevants features from a window of a frame in a video and , from this, train a support vector machine to classify human in a image by its descriptors based on gradient information contained in the histogram.

These techniques and methods was used in full exploitation of its benefits and good performance to the implementation of the pedestrian detection program.

## III. PROPOSED SOLUTIONS

In the first approach of this problem, the process of multi-frame-differencing with the adaptive background model as adopted to track moving regions and objects, as proposed by Xin Zhang *et al* [1]. However, this process was simplified to result in a more understandable and less computational costing process that was combined with morphology operations to remove possible noises and form in a more connected way the moving regions. The multi-frame-differencing was implemented in the algorithm as proposed, resulting frames less noised and with less false-positive regions.

At the process of adaptative background model, a built-in function from "OpenCV" was used to reproduce the process of subtraction of the scene from most of the stationary objects from the scene, and with the foreground mask generated (phenomenon described in the Figure 1), it is multiplied with the moving region obtained from the Multi-frame-differencing , producing a frame that contains only the objects belong to the foreground of the image and only regions where were registered movement, that is in fact significantly moving.



Fig. 1. An pedestrian moving in the scene

Obtained the moving objects from the scene, the HOG(histogram of oriented gradients) is performed in each frame from the video by an built-in function and, along with a pre-trained SVM(support vector machine) to classify human, the moving objects whose shape matches with the human shape are candidates of tracking in the process of detect pedestrian. This process is possible with the comparison of the descriptors that the HOG provides from a window of the image(process repeated along the entire frame and for each scene) with the information that the SVM learned to classify shapes.

Once its model provide a set of the frame's regions that SVM classify as human, its possible that in the same region of a pedestrian more than one subregion is detected as human shape, and as consequence, more than one rectangle detection is marked around the person( a event observed in Figure 2).



Fig. 2. More than one tracking in one person

For correct this kind of error, a nonmax suppression built-in function was used for to reduce this redundancy tracking error. This algorithm basically get every bounding boxes that surrounds human shape compare its ratio of overlapping between then, and being a thresh ratio set, every boxes that share a ratio bigger than the overlap thresh is supressed from

the region. Therefore, only sufficient distant bounding boxes will appear in the pedestrian detection.

The previous technique was implemented in the place of the analysis of the proportion of the moving objects in relation to the average human one that is given by the literature. One of the reasons of this replacement is that the idea tested to the extraction of the moving elements that had as base the manipulation of these elements through morphological operations to create bigger and filled connected elements, so that a algorithm of labelling connected objects could offer the solution needed. However, the complexity of the scene wouldn't enable this strategy, since a lot of components of the same pedestrian will remain unconnected.

After the identification of a possible human in the image, it is proposed by Xin Zhang *et al* [1] to make a classification of the figure based on the color of the skin of a human being and another classification based on the number of circles found in the region to determine the existence of a human head. About the algorithm proposed for the skin classifier, because of the lack of information about why the formula is created and the way that it must be implemented, together with the difficulty of determine the constant required, was decided to only verify the existence of a certain number o pixels in the moving object that satisfies the conditions given by S. Kolkur *et al* [5] so it can be classified as human skin using the HSV and BGR color spaces, and in consequence, a human as a whole.

Although the identification of a human by its skin color being simplistic and not very precise since a lot of things can have a color that is similar to a human skin and that it's very likely that some pixels in the region will pass the requirements of the classification, the algorithm still useful in case there is a monochromatic element or a low image resolution. Follow the conditions used:

$$0 \leq H \leq 50$$

$$0.23 \leq S \leq 0.68$$

$$R > 95, B > 20, G > 40$$

$$\max(RGB) = R$$

$$|R - G| > 15$$

In relation to the head identifier algorithm, it was used the model proposed with some alterations to better fit the classification through a histogram of oriented gradients(HOG). The head size was computed using the given formula and the circles were computed using the Hough Transform providing pre-defined possible head radius and using that to set a range to the radius of the circles since it gave a better result than not informing that information. However, because the rectangle size is controlled by the HOG analyser and do not follow the pattern of rectangle proportion exemplified, the lower boundary of the number of circles found was lowered by one to get better results without detecting a lot more of false-positive regions but resulting, in some situation, in more false-negative regions, since these regions aren't always completely connected as a consequence of the method used.

$$Headproportion = 0.0065 + -0.005$$

#### IV. EXPERIMENTAL RESULTS

We evaluate the above approach on real-life data. The camera is fixed with minimal movement at different distances from the pedestrian, close and far shots, losing some of the perspective of the person shape when she is big or too small for the camera. In addition to that were considered different ambient conditions together with different angles of capture. To realize the testing of the program were utilized around 100 frames from each condition video. Several example images from the training and test sets used are shown in Fig. 3. Image (a) is taken at 11:30 am in a sunny day while the image (b) is taken under the shadows. The multi-frame-differencing can efficiently detect the movement of objects with some consistency in images such as pedestrians and cars, presenting a confusion with plants, when not in a background with the same color then them. Using the method of HOG and SVM the algorithm is capable of ignoring small movements in the image and with the human head detection is capable of ignoring a lot of other moving objects of the scene, like you can see in image (c) and (d). Because also of the use of HOG, SVM and multi-frame-differencing the algorithm detects well people walking perpendicularly to the camera and close to it, but has problems identifying changes when the person is parallel to the camera or people far from it. The program is also capable of identifying a human even without all the body being in the frame and in different distances as is shown in the image (e) and (a). Using those test frames of 7 different videos, we computed the following analysis of the algorithm:

#### V. CONCLUSION

#### REFERENCES

- [1] Xin Zhang, Yuehua Gao, Xiaotao Wang, Jianing Li, Bing Wang, *Method for Detecting Pedestrians in Video*, 2012 International Conference on Systems and Informatics (ICSAI 2012)
- [2] P. KaewTraKulPong and R. Bowden, *An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection*, In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, AVBS01, Sept 2001.
- [3] N. Dalal and B. Triggs, *Histograms of Oriented Gradients for Human Detection*, INRIA Rhône-Alps, 655 avenue de l'Europe, Montbonnot 38334, France
- [4] CHRISTOPHER J.C. BURGESS, *A Tutorial on Support Vector Machines for Pattern Recognition*
- [5] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia, *Human Skin Detection Using RGB, HSV and YCbCr Color Models*, Advances in Intelligent Systems Research, 2017
- [6]
- [7]