STATIONARY FOREGROUND DETECTION USING BACKGROUND SUBTRACTION AND TEMPORAL DIFFERENCE IN VIDEO SURVEILLANCE*

Álvaro Bayona, Juan C. SanMiguel, José M. Martínez

Video Processing and Understanding Lab
Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain
E-mail: {Alvaro.Bayona, JuanCarlos.SanMiguel, JoseM.Martinez}@uam.es

ABSTRACT

In this paper we describe a new algorithm focused on obtaining stationary foreground regions, which is useful for applications like the detection of abandoned/stolen objects and parked vehicles. Firstly, a sub-sampling scheme based on background subtraction techniques is implemented to obtain stationary foreground regions. Secondly, some modifications are introduced on this base algorithm with the purpose of reducing the amount of stationary foreground detected. Finally, we evaluate the proposed algorithm and compare results with the base algorithm using video surveillance sequences from PETS 2006, PETS 2007 and I-LIDS for AVSS 2007 datasets. Experimental results show that the proposed algorithm increases the detection of stationary foreground regions as compared to the base algorithm.

Index Terms— Stationary foreground detection, background subtraction, frame difference, video surveillance.

1. INTRODUCTION

Intelligent and automated security surveillance systems have become an active research area in recent time due to an increasing demand for such systems in public areas such as airports, underground stations and mass events [1]. In this context, detection of stationary foreground regions is one of the most critical requirements for surveillance systems based on the detection of abandoned or stolen objects [2], or parked vehicles.

Background subtraction techniques are the most popular choice to detect stationary foreground objects [3][4][5][6], because they work reasonably well when the camera is stationary and the change in ambient lighting is gradual, and they represent the most popular choice to separate foreground objects from the current frame.

As suggested by [7], the existing approaches can be divided into two categories. The first category groups approaches that only use one background model, emphasizing those based on sub-sampling [4], where the selection of the sub-sampling rate is a critical parameter to detect stationary foreground objects, and those which analyse the input video frame by frame using techniques such as the accumulation of foreground masks [6] or techniques based on the properties of the background subtraction model used [5]. In [6], authors present a robust method that updates an intermediate image with foreground areas determined for every frame. Then, it is thresholded to detect stationary foreground. In [5], a method based on observing the transitions between different states in a background subtraction model based on GMM is presented. The second category groups approaches based on two or more background models, where the background models are analysed at different frame rates. The typical approach in this category [3] is based on two background models, where first model is updated every frame, trying to identify short-term changes, and the other model is updated every n frames, trying to find long-term changes. A recent study shows that sub-sampling approaches [4] obtain the best results because most false positives are removed due to the use of foreground masks from different instant times and by applying a logical AND combination stage.

In this paper, we extend the work presented in [4](from now on base algorithm). Firstly, the base algorithm is described. Then, some modifications are included to increase the robustness of the base algorithm by reducing false positives and handling occlusions. Finally, we have tested the base algorithm and its improvements with video surveillance sequences in two typical scenarios: parked vehicles and the abandoned or stolen object detection

The paper is structured as follows: section 2 overviews the base algorithm [4], section 3 describes the improvements added to the base algorithm, section 4 shows experimental results and section 5 closes the paper with some conclusions.

*Work supported by the Spanish Government (TEC2007- 65400 SemanticVideo), by Cátedra Infoglobal-UAM for "Nuevas Tecnologías de video aplicadas a la seguridad", by the Consejería de Educación of the Comunidad de Madrid and by the European Social Fund.

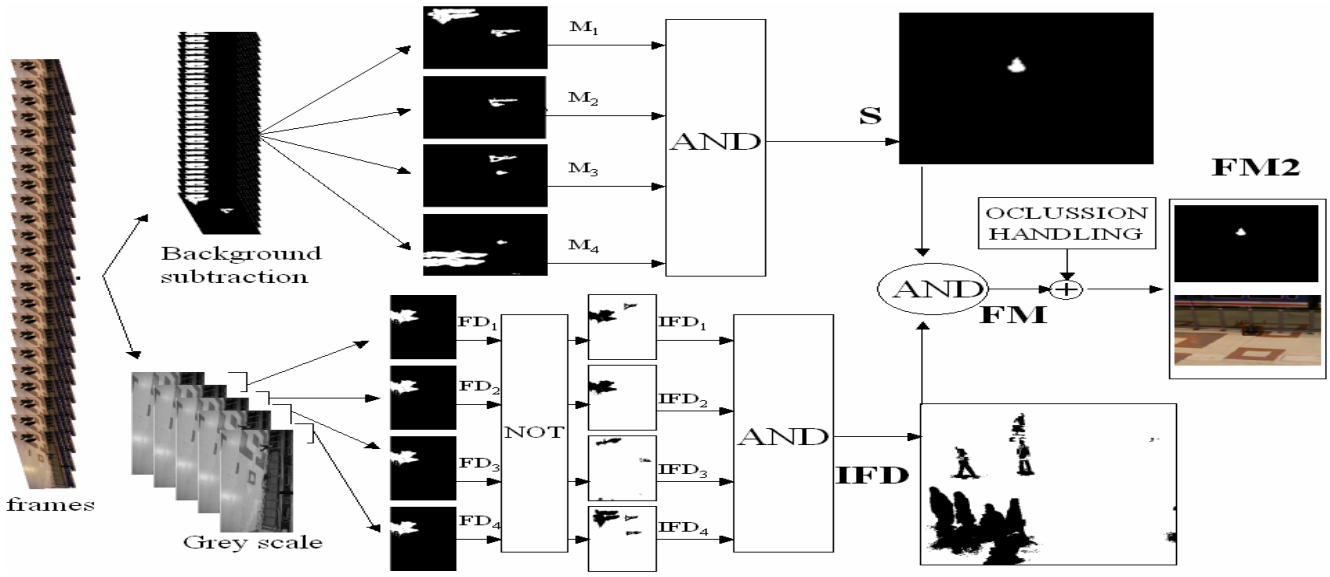


Figure 1: Proposed algorithm for stationary foreground region detection

2. BASE ALGORITHM FOR STATIONARY FOREGROUND REGION DETECTION

The stationary foreground region detection algorithm described in [4], is based on the sub-sampling of the foreground-mask with the aim of detecting foreground changes at different time instants in the same pixel locations. To achieve this, authors use a background subtraction stage based on modelling each pixel with a simple Gaussian distribution. Since it is assumed that the pixels of a stationary region will remain as foreground for a period of time, a number of binary foreground mask samples are collected from the last *k* seconds. Then, the stationary foreground mask (*S*) is obtained computing the intersection of the binary foreground mask samples. Finally, each active pixel of *S* is determined to be part of the stationary foreground regions. Specifically, the authors use 6 samples taken from the last 30 seconds. *S* is defined as follows:

$$S = M_1 * M_2 * M_3 * M_4 * M_5 * M_6 \qquad (1)$$

where $\{M_1 .... M_6\}$ are the foreground mask samples.

3. PROPOSED MODIFICATIONS

In this section, we describe some modifications introduced to improve the algorithm performance in many situations in terms of accuracy and computational cost. Figure 1 shows how, simultaneously, we obtain a mask *S* which studies the persistence of the foreground mask, and another mask *IFD* that shows the movement in the scene at the same time instants; both are combined for getting final masks FM and FM2 (see Fig.1).

3.1. Change of the background subtraction technique

The base algorithm described in [4] presents a high false positive rate due to the use of a simple Gaussian for the background subtraction stage. Simple Gaussian is a non-robust background subtraction approach because it is based on modelling isolated pixels. We propose to include a new background subtraction stage to solve the previously described problem. It is based on modelling each pixel considering its neighbours as described in [8]. This algorithm works on grey-scale images with static cameras and it is based on subtracting a square window around every pixel. The final decision is taken by thresholding the previously computed subtraction, also considering the noise introduced by the camera. Additionally, some modifications have been introduced to improve the processing time and to support background initialisation with moving objects.

3.2. Removal of false positives in crowded sequences

The base algorithm introduces several false positives in crowded sequences. In this context, moving people are always crossing certain regions of the scene and they might produce false stationary foreground regions in mask *S*.

To solve it, we propose to analyse the movement in these areas in order to remove false stationary foreground pixels from *S* (the ones belonging to moving regions), but not stationary regions. This is achieved by using a frame difference technique. As demonstrated for the background subtraction stage, we decided to apply a sub-sampling frame difference scheme. For each sampling instant, we compute the grey scale difference between the current frame and the frame of the previous sampling instant. Then, the difference is thresholded to obtain a motion mask (FD_K). After that, we make a logical inversion of the motion masks to obtain non-

motion masks (IFD_K). In the next step, the final non-motion mask (*IFD*) is computed by applying a logical AND between all the non-motion masks. Finally, the stationary foreground region mask *FM* is obtained by applying a logical AND between the final non-motion mask (*IFD*) and the stationary foreground mask obtained from the background subtraction stage.

With respect to the parameters, we have used the same number of stages proposed in the base algorithm, so we obtain 6 frame difference masks for computing the final non-motion mask (*IFD*).

3.3. Tolerance to occlusions

After testing the second modification, we observed that the algorithm did not support partial and total occlusions (not like the base algorithm, which supports occlusions). This is because the IFD mask sets occluded pixels to '0' when a moving person crosses into the camera and occludes the stationary object. In order to solve this problem and, therefore, to increase the robustness of the algorithm in complex sequences (where these occlusions take place many times), we have included a modification to detect occlusions and correct them.

Firstly, we extract and keep information about blobs from the FM mask obtained in the last sampling stages. Then, a comparison between current detected blobs and the previously detected blobs (in the preceding sampling instant) is performed. If any blob is missing in the current list, we study if it is because the stationary region has disappeared or if there is an occlusion. We detect an occlusion if there is a moving region detected in the background subtraction stage (pixel value of *S* mask equals to '1') and the frame difference stage (pixel value of IFD equals to '0').

If an occlusion is detected, we select the bounding box corresponding to the occluded region from the *S* mask, and calculate the percentage of active pixels. If this percentage is over a threshold τ, we consider that the stationary object has been occluded, so we copy the corresponding blob, stored in memory, to the current FM mask, and include this blob in the current blob list. Therefore, the final static foreground object mask is obtained by joining occluded objects masks with the FM mask into a new mask FM2.

Dataset	View	#Seq	Complex Level	Noise Level	Occlusions	Static Regions
PETS 2007	Cam 1	5	Very High	High	Very High	7
	Cam 2	5	Very High	Medium	Very High	10
	Cam 3	5	Medium	High	Low	9
	Cam 4	5	High	Low	High	12
AVSS 2007	AB	3	Medium	Low	Medium	7
	PV	4	High	High	Low	8
PETS 2006	Cam 1	7	Medium	High	Low	14
	Cam 3	7	Very Low	Low	Very Low	8
	Cam 4	7	Medium	Low	Medium	13

Table 1: Description of properties of each test sequence

4. EXPERIMENTAL RESULTS

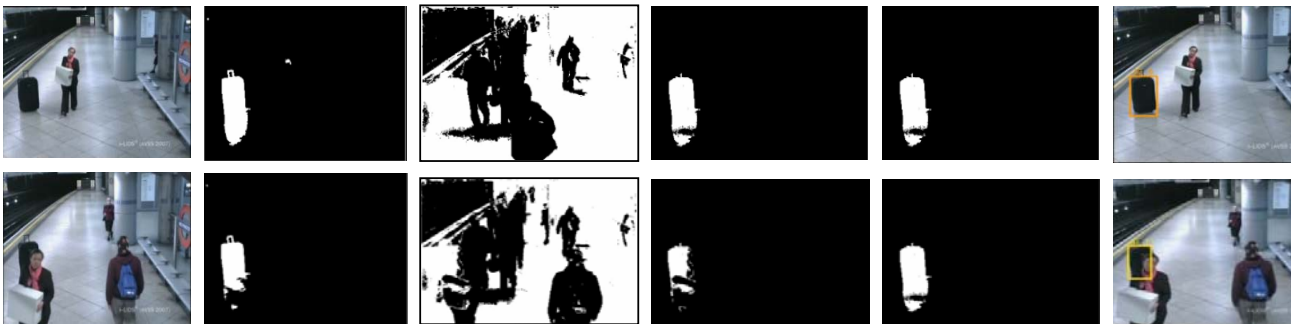
In this section, experimental results of the base algorithm and of the proposed enhancements are presented. The system has been implemented in C++, using the OpenCV image processing library (available at http://sourceforge.net/projects/opencv/). Tests were executed on a Pentium IV with a CPU frequency of 2.8 GHz and 1GB RAM.

Experiments were carried out on selected sequences (see Table 1) from the i-LIDS dataset for AVSS2007 (available at http://www.avss2007.org), the PETS2006 dataset (available at http://www.pets2006.net/), and the PETS2007 dataset (available at http://pets2007.net/). Table 1 shows a summary of the main properties of the selected sequences. We have decided to classify these sequences depending on their characteristics, which are the difficulty of extracting stationary foreground regions, the noise introduced into the sequence by the recording device and the number of occlusions. We have ranged these characteristics between very high and very low.

The proposed algorithm has two critical parameters, the number of sub-sampling stages and the sampling frequency. Initially, we have considered the values proposed in [4] (6 stages, 30secs) and for the simple sequences, the results are quite satisfactory, but for complex sequences, results get significantly worst. In order to compare our proposal against

	PETS 2007								AVSS 2007				PETS 2006					
	Camera1		Camera2		Camera3		Camera4		AB_Data		PV_Data		Camera1		Camera3		Camera4	
	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R
Base algorithm [4]	0,08	0,57	0,03	0,10	0,12	0,33	0,04	0,33	0,38	0,57	0,08	0,25	0,32	0,54	0,64	0,75	0,37	0,57
+ Section 3.1 enhancements	0,19	0,87	0,12	0,70	0,27	0,77	0,28	0,75	0,6	1	0,27	0,62	0,54	1	0,88	1	0,65	0,92
+ Section 3.2 enhancements	0,51	0,57	0,55	0,30	0,42	0,66	0,62	0,66	0,75	0,71	0,45	0,87	0,75	0,81	1	1	0,85	0,85
+ Section 3.3 enhancements	0,53	0,87	0,57	0,40	0,45	0,77	0,65	0,75	0,81	1	0,45	0,87	0,84	1	1	1	0,87	0,92

Table 2: Comparative results for each group (in percentage)



Current frame *S Mask (3.1)* *IFD Mask* *FM Mask (3.2)* *FM2 Mask (3.3)* *Results*

Figure 2: Comparative results for frame 4554 and 4630 (next sub-sampling frame) of AB_easy

the base algorithm, we keep the 30s value for detecting a foreground region as stationary. We have tested a number of stages between 2 and 30, observing that, for crowded sequences the selection of few sampling stages increases the false positives, whilst when using many sampling stages, the stationary region could be occluded and never detected. Finally, we have empirically chosen to use 4 sampling stages, with a sampling frequency of 170 frames (24-25fps).

To evaluate and compare the performance of the proposed algorithm, we have manually annotated the stationary regions to obtain a ground truth. The experimental results obtained with the base and proposed algorithms are summarized in terms of precision and recall in Table 2 for the stationary region detection. Table 2 shows how each modification introduced in the base algorithm improves gradually the results obtained for precision and recall. The base algorithm detects stationary foreground regions with high accuracy in the simple sequences like camera 3 from PETS2006 but it introduces some false positives into the final mask. For the complex sequences, its accuracy is dramatically decreased because of the high density of moving people. Then the background subtraction technique is changed (section 3.1). It removes isolated false positives, so precision and recall are improved for all sequences. However, the accuracy in complex sequences is still low. Then, the frame difference stage is integrated (section 3.2). Precision in simple and complex sequences is increased but Recall is maintained in simple sequences and slightly reduced in complex sequences due to occlusions. Finally, the modification to handle occlusions is introduced (section 3.3). The Recall measure is heavily improved and the occluded regions are correctly detected. Figure 2 shows two examples of the proposed algorithm. The first and second rows show, respectively, a stationary foreground region detection without occlusions and under a partial occlusion. It can be noticed that the FM mask does not contain the entire stationary region (the same as in the first row) because a person is occluding the region. The occlusion is detected and corrected as shown by the FM2 mask.

5. CONCLUSIONS

This paper has presented an algorithm for stationary foreground region detection. As base algorithm [4] we selected one based on a sub-sampling scheme, due to its better performance [7]. Then, some improvements have been included to solve the main drawback of the base algorithm: its high false positive rate in crowed sequences. A change of the background subtraction technique, an integration of a sub-sampled frame difference stage and an occlusion management mechanism have been proposed to reduce this problem. Experimental results show that the proposed modifications improve the results obtained by the base algorithm, reducing the false positive rate in crowed sequences.

6. REFERENCES

[1] Valera, M., Velastin, S.A. “Intelligent distributed surveillance systems: a review”. IEE Proceedings - Vision, Image and Signal Processing, 152(2):192-204, 2005

[2] Ferrando, S.; Gera, G.; Regazzoni, C. “Classification of Unattended and Stolen Objects in Video-Surveillance System”. Proc. of AVSS 2006, pp. 21-27.

[3] Porikli, F.; Ivanov, Y.; Haga, T. “Robust Abandoned Object Detection Using Dual Foregrounds”, Journal on Advances in Signal Processing, vol. 2008, art. 30, 11 pp.

[4] Liao,H-H.; Chang,J-Y.; Chen, L-G. “A localized Approach to abandoned luggage detection with Foreground –Mask sampling”, Proc. of AVSS 2008, pp. 132-139.

[5] Mathew, R.; Yu, Z.; Zhang, J. “Detecting new stable objects in surveillance video”, Proc. of MSP 2005, pp. 1-4

[6] Guler, S.; Farrow, K. “Abandoned Object detection in crowded places”, Proc. of PETS 2006, June 18-23.

[7] Bayona, A.; San Miguel, J.C; Martínez, J.M: "Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques", Proc. of AVSS2009, pp. 25-30

[8] Cavallaro,A., Steiger O., Ebrahimi T., “Semantic Vídeo Analysis for Adaptive Content Delivery and Automatic Description”, IEEE Transactions of Circuits and Systems for Video Technology, 15(10):1200-1209, October 2005.