

From Moving edges to Moving Regions

L. Biancardini¹, E. Dokladalova¹, S. Beucher², and L. Letellier¹

¹ CEA LIST

Image and Embedded Computer Laboratory,
91191 Gif/Yvette Cedex, France

² Centre de Morphologie Mathématique ENSMP,
35 rue Saint Honoré 77305 Fontainebleau cedex, France
¹ {biancardini, eva.dokladalova, letellier}@cea.fr
² {beucher}@cmm.ensmp.fr

Abstract. In this paper, we propose a new method to extract moving objects from a video stream without any motion estimation. The objective is to obtain a method robust to noise, large motions and ghost phenomena. Our approach consists in a frame differencing strategy combined with a hierarchical segmentation approach. First, we propose to extract moving edges with a new robust difference scheme, based on the spatial gradient. In the second stage, the moving regions are extracted from previously detected moving edges by using a hierarchical segmentation. The obtained moving objects description is represented as an adjacency graph. The method is validated on real sequences in the context of video-surveillance, assuming a static camera hypothesis.

1 Introduction

Automated video surveillance applications have recently emerged as an important research topic in the vision community. In this context, the monitoring system requirement is to recognize interesting behaviors and scenarios. However, in such a system, the main problem is to localize objects of interest in the scene. In this context, every moving area is potentially a good region of interest.

There are three conventional approaches to automated moving target detection: background subtraction [5–7, 13], optical flow [5, 8] and temporal frame differencing [5, 10, 14]. In video surveillance, the background subtraction is the most commonly used technique. However it is extremely sensitive to dynamic change of lighting. Nevertheless, it requires a prior knowledge of the background, which is not always available. In the second category of methods, the optical flow estimation is used as a basis for further detection of moving objects. However, it is a time consuming task. It is affected by large displacements and does not provide the accurate values, neither at moving objects contours, nor in large homogeneous areas.

In this paper, we focus on the temporal frame differencing methods. These techniques enable fast strategies to recover moving objects. However, they generally fail to extract accurately both slow and fast moving objects at the same time. In such case, a tradeoff between missed targets and false detections is very

hard to obtain. To overcome these problems, we first propose a new difference scheme suited to moving objects boundaries detection. Then, a hierarchical segmentation [1–3] of the current frame is used to complete these contours and extract the underlying moving regions.

The paper is organized as follows: section 2 introduces the method for motion boundaries extraction. In section 3, the use of the hierarchical segmentation to retrieve the moving regions is described. The experimental results are presented in section 4. Then, we give the conclusions on the proposed method and we discuss the future work.

2 Moving Edges Detection

The frame differencing methods take advantage of occlusions, which occur at moving objects boundaries. Various kinds of approaches have been attempted in the literature [10, 11, 13, 14]. Generally, the presence of the occlusions is detected using the absolute difference of two successive frames. However, the occlusions do not correspond to the position of the true object boundaries neither in the first image nor in the second one. Moreover, depending on frame rate and speed of the moving objects, the difference map can critically differ. When an object moves slowly, image intensities do not change significantly in its interior. Consequently, the resulting difference image exhibits high values only at motion boundaries. In the opposite case, if the object has completely moved from its position, the resulting frame difference will exhibit high values inside the object body in both images. It is the so-called ghost phenomena [12] and leads to false detections.

In [13, 14], the authors propose to use a double-difference operator. The frame difference is performed on the two pairs of successive images at time $(t-1, t)$ and $(t, t+1)$. Then, the result is obtained by the intersection of these two difference maps. However, when an object moves slowly this intersection may be reduced to an insufficient number of pixels.

2.1 Difference scheme

In the following, $I^t : \mathbb{Z}^2 \rightarrow \mathbb{N}$ indicates a discrete image at a given time $t \in (0, T]$. We note the reference frame, the frame in which we want to localize and segment the objects in motion.

The proposed method considers three successive images I^{t-1}, I^t and I^{t+1} . We assume that moving edges position depends rather on the gradient changes in the successive images than in the images themselves.

First, we compute the spatial gradient modulus of I^{t-1}, I^t and I^{t+1} and we note $g^t = \|\nabla I^t\|$ (respectively $g^{t\pm 1}$). Then, the symmetrical frame difference is obtained on the two pairs of gradient images. The moving edges measurement at a given time t is defined as the infimum operator of the two difference maps:

$$mem^t = \inf(|g^{t+1} - g^t|, |g^t - g^{t-1}|) \quad (1)$$

The infimum operator properties and the analysis of gradient over three frames yield the interesting behaviors:

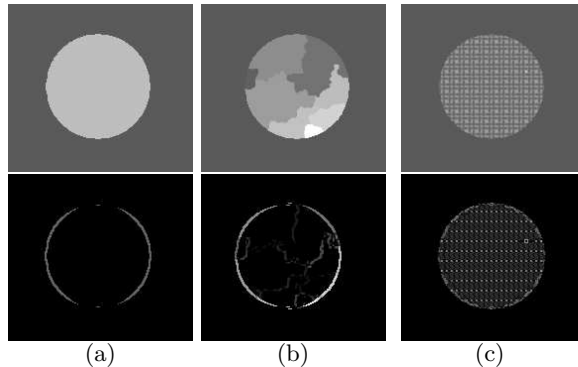


Fig. 1. Results of *mem* on three different cases: (a) homogeneous region, (b) assembly of homogeneous regions and (c) textured area.

1. a maximum response at moving objects boundaries locations: when a contrasted object is moving over a homogeneous area, the *mem* is equal to the original gradient in the reference frame.
2. a significant robustness to motion amplitude: in the case of fast moving objects, the result is not delocalized and the ghost phenomena are drastically reduced.
3. a significant robustness to random noise (non-repeatable in subsequent frames).

However, due to low motion, weak contrast with the scene and the aperture problem (sliding contours), the moving edges measurement will certainly fail to provide information along the whole contours of a moving object (figure 1). The forthcoming section explains how to overcome this problem in order to obtain reliable moving regions.

3 From Moving Edges to Moving Regions

In this section, we propose a method to extract moving regions based on the new moving edges measurement (*mem*) proposed in paragraph 2.1. However, the *mem* operator does not result in the complete object contours. Thus we propose to consider an additional information issued from a spatial segmentation of the reference image. Nevertheless the segmentation process generally results in an over segmentation of the image, an accurate description of the image requires multiple levels of details. Thus, in our approach, the moving regions are searched through the levels of a hierarchical segmentation, which allows to study the regions at different scales.

We start by extracting an initial set of moving contours corresponding to spatial edges with a sufficient *mem* value. Then, the moving objects are detected by browsing a set of candidate regions extracted from a hierarchical partition.

3.1 Hierarchical segmentation and candidate set to detection

Some attempts to extract the meaningful image regions by gathering the regions of an initial segmentation can be found in [2, 1, 4]. However, they are not based

on exhaustive analysis of region grouping which have a significant computational complexity. As explained in these publications one way to reduce the number of candidates is to build a hierarchical segmentation. After an initial partition is built, a graph is defined, by creating a node for each region and an edge for each adjacent regions pair. Graph's edges are weighted according to a dissimilarity criterion (as example, a grey level difference) between two regions. The hierarchical segmentation is obtained by progressively merging regions of the initial segmentation, in an increasing dissimilarity order. The process is iteratively repeated until only one region remains. By keeping track of the merging process, we construct the candidate set of regions C . Each time two regions merge, the resulting region is added to the candidate list. Note that the candidates are sorted according to their level of apparition in the hierarchy. The total amount of distinct regions in the candidate list is $2N-1$ (N is the number of regions of the initial partition) [1]. This hierarchical segmentation only contains the more meaningful assembly of regions in the sense of the chosen dissimilarity criterion.

In our approach, we use the set of contours and regions given by the watershed transform proposed in [3]. We choose a robust dissimilarity criterion based on the contrast: for a given regions pair, the value of the criterion is defined by the median value of image gradient modulus along the watershed lines separating the regions.

3.2 Initialization Step: Extraction of moving contours

Once the hierarchical segmentation is built, the next step of the algorithm is to extract a set of moving contours: the *mem* is calculated and a threshold is applied to obtain a binary image. A set of moving points designed as the most significative contours in motion (*mscm*) is obtained by intersecting the thresholded *mem* image with the lowest level's contours in the hierarchy. The resulting binary image (section 4, figure 3(b)) may not contain the whole moving object's boundary but only some incomplete and fragmented parts. Consequently, the next step of our method is to gather and complete moving contours coming from a same object, and discard small or isolated components corresponding to residual noise. True moving edges are supposed to be distributed with enough coherence and density around a same region to be gathered as the contours of this region. *A contrario*, noise components are sparse and dispersed. They can not be assembled as the contours of any region in the hierarchy.

3.3 Detection of moving regions

The detection step is achieved by independently optimizing a local criterion on each region of the candidates list defined in section 3.1. In the following, for any given candidate $C_i \in C$, the frontier ∂C_i of the region is defined as the subset of watershed points enclosing C_i . The matching score of a region C_i is calculated as the proportion of significative contours in motion contained in its frontier ∂C_i . This is simply expressed by:

$$ms(C_i) = \text{card}(\partial C_i \cap \text{mscm}) / \text{card}(\partial C_i) \quad (2)$$

where card refers to the cardinal operator.

Each candidate is successively tested according to its order of apparition in C . A candidate is labeled as detected if its score $ms(C_i)$ is higher than a predefined threshold $T_{percent} \in [0, 1]$.

The method may lead to some incorrect detections as depicted in figure 2(c). In figure 2(c), the region C_2 which causes the error is detected because its frontier in common with the moving region C_1 (figure 2(b)) is quite long. Nevertheless, the frontier of C_2 is longer than the frontier of C_1 but does not contain more moving contours points (figure 2(a)). Consequently, the score $ms(C_i)$ of the region C_1 is higher than the one of C_2 which enables to select the final correct region.

As explained in section 3.1, each candidate (except those from the initial partition) comes from a merging sequence of some preceding candidates. Owing to the construction of C , when C_i is tested as a moving region, it implies that all the grouping candidates constituting this sequence are already processed and their scores are known. In the following, we will refer to this set of candidates as the set of Ancestors of C_i . To avoid situations such as depicted in figure 2, we propose to add the following condition to the detection: If C_i exhibits a score superior to $T_{percent}$, C_i is said to be detected if and only if, its score is higher than any score of its ancestor. This can be expressed by adding the following condition:

$$ms(C_i) > \max_{C_k \in \text{Ancestors}(C_i)} (ms(C_k)) \quad (3)$$

This is a sufficient way to discard many false detections. However, when candidate regions correspond to higher levels of the hierarchy, their frontiers are longer. Thus, a significant score is more difficult to obtain (a larger portion of the contour may be missing). In that case, the score can not be constrained to be strictly higher than the previous ones. Consequently, we propose to test whether the new matching score is higher than the previous ones multiplied by a weighting coefficient α , taken in the interval $[0,1]$.

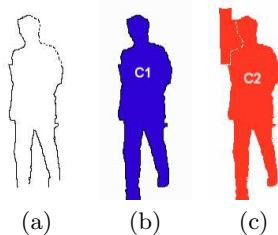


Fig. 2. (a) the $(mscm)$ set (b) a first candidate matching the contours (c) a later candidate matching the contours

4 Experimental Results

In this section, we present some results on video sequences corresponding to real situations of video-surveillance.

In the presented results, we use the regularized Deriche gradient [9] to obtain the moving edges measurement (see section 2). The regularization parameter σ should be chosen greater than 2.0 in order to preserve poorly contrasted or narrowed structures.

The threshold parameter T_{mem} used to obtain the m_{scm} mainly depends on the level of noise in the mem image. Nevertheless the experiments show that it is stable over time for a given scene and a fixed video camera. In all the presented experiments this parameters is set $T_{mem} = 2.0$. The result of m_{scm} detection is presented in figure 3(b).

As it was presented in the section 3, we use the watershed transform to obtain the initial segmentation. In order to reduce (once again) the computational cost of the algorithm we propose to use a reduced set of markers. It is obtained by the h-minima operator with $h = 3$ [15]. During the detection process, we use $T_{percent}$ to express the ratio of the target boundary length, which can be missing without altering the detection of the corresponding region. This parameter was set to 0.65, which enables to detect the regions from an incomplete set of moving contours, without generating false detections. The experimentally verified best values range of parameter α is [0.65, 0.85]. The $alpha$ paramater's influence can be reduced by taking into account the size ratio of the currently treated region and its detected ancestors in the algorithm of section 3.3.

The initial set of moving edges is presented in the figure 3(b). The contours of all the regions detected during the matching process are shown in figure 3(c). Once the detection is achieved, isolated components with area under 50 pixels are removed and the remaining regions are merged according to the dissimilarity criterion. The results of this post-processing step are shown in figures 4(a) and 4(b).

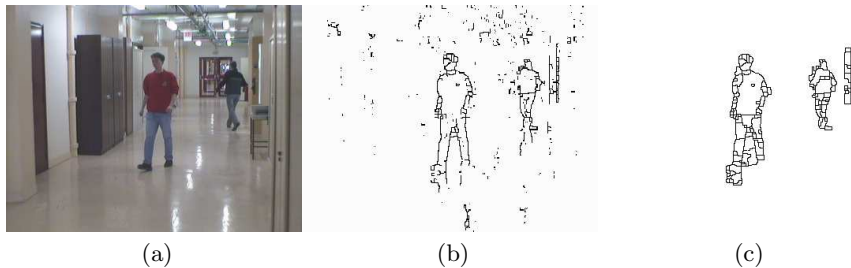


Fig. 3. (a) original image, (b) set of moving contours initially detected (c) contours of detected regions after parsing the hierarchy.

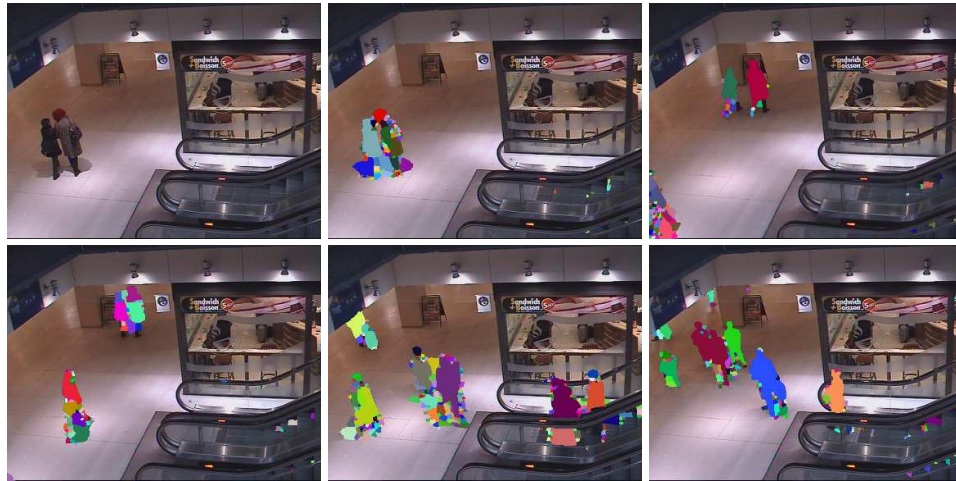
5 Conclusions

This paper focuses on the extraction of moving objects in the video-surveillance context. The goal is to detect all potential zones of interest and create their representation suitable for tracking and scene interpretation.

First, we introduce a new method to perform the detection of moving objects boundaries. The moving edges are extracted with an operator based on the double differences of three successive gradient images. The defined operator is



(a)



(b)

Fig. 4. From top-left to bottom-right (for each data set): first image of the sequence and moving regions detected in some subsequent frames.

robust to random noise and the results are not affected by the displacement speed of objects. Then we show how to use the hierarchical segmentation in order to pass efficiently from the incomplete detected contours to the entire regions in motion. To obtain the accurate set of moving regions, we propose to combine two criteria during the detection process: i) the contrast criterion ii) the matching score criterion. The hierarchical approach also reduces the computation time that is the limiting factor in the video-surveillance applications.

Another advantage of the method is that the extraction of the moving objects requires neither motion calculation nor prior knowledge of the scene.

In addition, the moving targets are extracted as an assembly of multiple homogeneous parts of different size and contrast. Due to the underlying hierar-

chical segmentation structure, their adjacency and inclusion relations are known. These considerations are very useful to construct a model for the detected targets. This model can be then used in several ulterior steps such as tracking, occlusions analysis or pattern recognition.

Consequently, the next stage of our work will concentrate on the study of the hierarchical graph-based object description for the scene interpretation and the object tracking in the security domain.

References

1. P. Salembier, L. Garrido. Binary Partition Tree as an Efficient Representation for Image Processing, Segmentation, and Information Retrieval, .IEEE Transactions on Image Processing, 9(4):561-576, April 2000.
2. F. Zanoguera, B. Marcotegui, F. Meyer, A Toolbox for Interactive Segmentation Based on Nested Partitions, ICIP, Kobe (Japan), 1999
3. Beucher, Segmentation d'images et morphologie mathématique, Doctorate thesis, Ecole des Mines de Paris, Cahiers du centre de Morphologie Mathématique, Fascicule n 10, Juin 1990.
4. Stan Sclaroff, Lifeng Liu: Deformable Shape Detection and Description via Model-Based Region Grouping. IEEE Trans. Pattern Anal. Mach. Intell. 23(5): 475-489 (2001)
5. D. S. Zhang, G. Lu, Segmentation of Moving Objects in Image Sequence: A Review, Circuits, Systems and Signal Processing (Special Issue on Multimedia Communication Services), 20(2):143-183, 2001.
6. Srinivas Andra, Omar Al-Kofahi, Richard J. Radke, and Badrinath Roysam, Image Change Detection Algorithm: A systematic Survey, Submitted to IEEE Transactions on Image Processing, July 2003.
7. M. Piccardi, Background subtraction techniques: a review, in Proc. of IEEE SMC 2004 International Conference on Systems, Man and Cybernetics, The Hague, The Netherlands, October 2004.
8. M.J. Black, P. Anandan, The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields, Computer Vision and Image Understanding, CVIU, 63(1), pp. 75-104, Jan. 1996.
9. R. Deriche, Fast algorithms for low-level vision, IEEE-PAMI, 12(1):78-87, 1990.
10. N. Paragios, R. Deriche, A PDE-based Level Set Approach for Detection and Tracking of Moving Objects, In Proceedings of the 6th International Conference on Computer Vision, Bombay, India, Jan. 1998.
11. Shi, Malik, Motion segmentation and tracking using normalized cuts, University of California, Berkeley report nUCB/CSD-97-962, 1997
12. R. Cucchiara, C. Grana, M. Piccardi, A. Prati, "Detecting Moving Objects, Ghosts and Shadows in Video Streams" in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, n. 10, pp. 1337-1342, 2003
13. K. Toyarea, J. Krumm, B. Brumitt, and B. Meyers, Wallflower: Principles and practice of background maintenance," in International Conference on Computer Vision, 1999, pp. 255-261.
14. Yoshinari, K. and Michihito, M., "A human motion estimation method us using 3-successive video frames Proc. of Intl. Conf. on Virtual Systems and Multimedia, 1996, pp. 135-140.
15. Tae Hyeon Kim, Young Shik Moon, "A New Flat Zone Filtering Using Morphological Reconstruction Based on the Size and Contrast," VLBV, 1999