REAL TIME VIDEO FOREGROUND EXTRACTION BASED ON CONTEXT-AWARE BACKGROUND SUBTRACTION*

Alvaro García, Jesús Bescós

Grupo de Tratamiento de Imágenes, Escuela Politécnica Superior
Universidad Autónoma de Madrid, E-28049 Madrid, Spain
e-mail: {Alvaro.Garcia, J.Bescos}@uam.es

ABSTRACT

This paper describes a real-time approach for foreground segmentation in image sequences taken from a stationary camera, as a first preprocessing step for object extraction. The proposed method combines a background subtraction method with the extensible inclusion of context information. It currently considers bidirectional temporal change detection as the a priori context data (i.e., desired objects are moving). This combination has proven to improve the detection of the object boundaries, hence enhancing the reliability of the segmentation masks, with little extra computational load. The tested implementation of the proposed foreground detector shows a good trade-off between efficiency, segmentation accuracy and robustness, particularly compared with some other state-of-the-art methods.

Index Terms**—** video analysis, background subtraction, change detection, context information.

1. INTRODUCTION

Foreground detection, or background extraction, is a very common preprocessing step for many higher-level applications in computer vision.

The majority of methods in the literature rely on variations of two basic approaches: frame difference evaluation and background subtraction. Some of them make use of both, trying to balance between their advantages and drawbacks.

For most surveillance applications, the camera is assumed to be stationary. In this context, methods based on background subtraction have obtained the best results. Many different techniques can be applied to classify foreground pixels and then construct a background model, like a mixture-of-gaussians model (GMM)[1][2][3], eigen-backgrounds[4], or mean-shift based estimation[5][6]. A brief review on these and other popular techniques can be found in[7].

In order to improve segmentation, other techniques are combined with background subtraction[8][9]. In [8], a running gaussian average is combined with temporal change detection in order to prevent errors in the background update stage. However, it is not used to improve directly the segmentation performed by background subtraction. This results in a slight improvement in the subtraction phase. As explained in the next section, temporal change can also be exploited in the foreground/background classification stage of the background subtraction method.

In [10] a novel method for moving object detection is proposed, combining bidirectional frame difference with background difference. It achieves good results with highly textured objects or when there are high differences between consecutive frames (e.g., for low frame rates). When applied to typical sequences, where objects can be low textured and there is little change between consecutive frames, homogeneous regions inside objects are incorrectly classified as background. However, areas belonging to the edges of moving objects (where changes are more obvious) are still well classified. The thickness of these borders increases when the objects position changes significantly from one frame to the next. Foreground objects detected by this method are usually not completely detected. However, it seems a good starting point for more sophisticated methods, especially taking into account the fact that most of the background subtraction approaches reduce the quality of their segmentation at low frame rates, frequently used for high resolution real-time operation.

Our goal is to classify, in real time, points on the image not belonging to the background. which for the targeted context is assumed to be non-complex. This means that there are no moving elements in the background. As this segmentation will be later used for object detection and tracking, we will give special importance to object boundaries in the foreground segmentation, as they will define object shapes.

This paper presents a method that combines a context driven model of the foreground to improve its detection via background subtraction. The presented approach applies bidirectional temporal change detection (i.e., desired objects are moving) to achieve it.

This paper is structured as follows: section 2 describes the proposed approach; section 3 presents a working application, and experimental results; finally, section 4 draws the conclusions.

* Work partially supported by the Spanish Government under Project TIN2004-07860-C02-01 (Medusa).

2. ALGORITHM DESCRIPTION

This method is designed to work with different *point* sizes, which can be either pixels or pixel blocks. Election of the size of these points will depend on the requirements of the application. Grouping pixels in blocks will provide greater efficiency and robustness against noise, although obtained mask will not have pixel-accuracy. Working with pixels will provide a finer segmentation, with loss of efficiency and robustness.

The working flow of the full method is illustrated in Figure 1. The first stage performs a temporal change detection, which will try to detect the moving foreground. This first segmentation is detailed in subsection 2.1. The mask obtained from this first stage will be used as context information in the second stage.

The second stage consists of a context-aware background subtraction algorithm, which will yield the final segmentation. It is a modified running gaussian average algorithm that takes into account not only the background model and the incoming frames, but also external context information, currently representing a priori confidence about moving objects. This algorithm will be described in section 2.2.

.

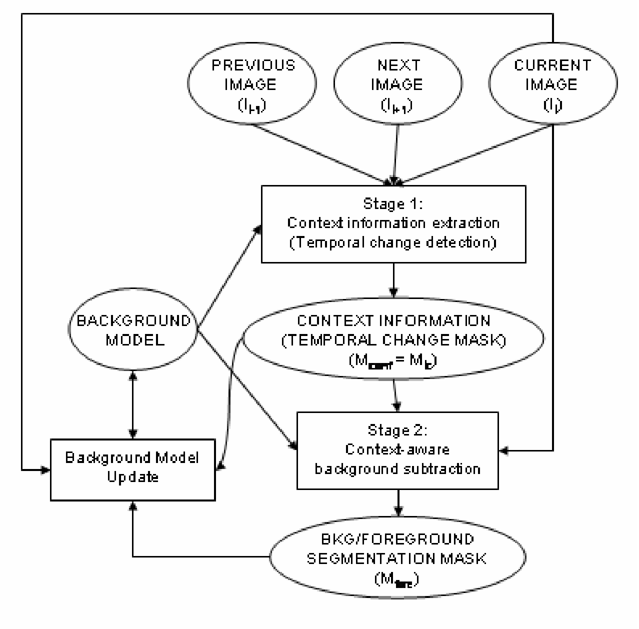

Figure 1*: Overview of the implemented method*

2.1. Temporal change detection

This stage is based on the method proposed in [10]. Temporal change is explored in both directions. The previous and posterior frames are compared to the considered frame obtaining two frame difference masks (called $M_{fd\text{-}prev}$ and $M_{fd\text{-}post}$) and a rough background subtraction is also performed (M_{bg_sub}). Obtained masks are combined in order to get an initial segmentation mask :

$$M_{tc} = (M_{fd\text{-}prev} \text{ AND } M_{bg\_sub}) \text{ OR } (M_{fd\text{-}post} \text{ AND } M_{bg\_sub})$$

This method is efficient and shows little false positives (background points classified as foreground), but fails in homogeneous (low-textured) areas of both the foreground and the non-moving objects, where points are very similar between consecutive frames, and frame difference is not able to robustly detect changes.

In order to improve robustness against noise and parameter values, we perform background subtraction with a single-gaussian model, instead of the fixed thresholding method used in [10]. The appearance of speckles in classified areas usually associated with gaussians in the background model is prevented by the combination with frame difference masks.

2.2. Context-aware background subtraction

Assuming a non-complex background in the input sequences (static camera, no moving elements in the background), a single gaussian model for each point is sufficient. As explained in [7], a mixture of gaussians algorithm with a single gaussian in its model is equivalent to a running average gaussian. In this way, we are modelling the averaged value of each point (mean value) and an estimation of its noise over time (the standard deviation).

Context information aims to provide confidence on each point's belonging to the foreground. In this sense, it conforms an a priori confidence mask, represented by M_{conf}. Currently, this mask results from the segmentation performed in the temporal change detection stage (i.e., $M_{conf}=M_{tc}$) . However, we are testing the inclusion of other context based criteria in the same framework. Luminance and texture homogeneity, object connectivity and compactness, or coherence of the object motion (extracted via a tracking algorithm [9]), could be easily integrated into the system, providing improvement in robustness and sensitivity.

Our algorithm is based on the one described in [8], modified in two ways in order to account for context information both in the classification and in the background model updating phases, as explained in the following subsections..

2.2.1. Variable weight background/foreground classification

The first modification is the use of M_{conf} in the background/foreground classification process. In the classic running gaussian average[7], a point value in I_t is classified as foreground if the inequality:

$$|I_t - \mu_t| > k \cdot \sigma_t$$

holds, being μ_t and σ_t the mean and standard deviation for that point in the background model, and k a fixed value. In our approach, we propose k to be dependent of the value of M_{conf} for each point. The higher the confidence in a point's belonging to foreground, the lower the value of k for that point.

As we currently use a binary version of M_{conf} (it represents presence or absence of motion), k will take one of two possible values, k_{min} for changed points and k_{max} for non-changed points.

The result of this classification will provide the final background/foreground segmentation mask (M_{fore}).

2.2.2. Robust background model adaptation

Currently, methods based on background difference update their background model only in those points classified as background. This prevents foreground objects from corrupting the background model. Tjis motivates our second modification to the basic algorithm. We combine the received contextual information with the final background/segmentation mask before using it for selective background update. In the current implementation, as contextual information is represented by a binary mask, they are just or'ed to obtain the background model update mask (M_{bkg_upd}). To ensure a clean background model, only points with a high degree of confidence about its belonging to the background (i.e. inside M_{bkg_upd}) will be updated in the background model, preventing typical foreground classification errors from polluting the background model.

This proposed method includes into the foreground concept all objects that are not initially in the background: moving objects, shadows, reflections, stopped objects that were not initially into the background and background regions initially occluded by static background objects that began to move. This approach reduces the risk of foreground objects being blindly included into the background, and allows for an object detection system that can later distinguish between them, applying its own knowledge, based on context information, of the kind of objects that can be present in the foreground.

Assuming a non-complex background, common in surveillance applications, static objects initially detected as foreground would not be blindly included into the background as time passes by. Instead, a maximum time is set for a static object to be consciously included into the background model.

3. EXPERIMENTS AND RESULTS

The described implementation has been tested into a working system: the Medusa project. The goal of one of its applications is to count people crossing an inter-building corridor with glass walls to the exterior. This scenario provides a non-complex background with high degree of illumination changes, shadows and reflections.

This application has been tested using different point sizes: 8x8 blocks and pixels. Working with 8x8 blocks greatly improves efficiency at the expenses of lowering shape resolution, but this might be enough for some applications as the one we show. Moreover, this could act as a rough confidence mask for a pixel level segmentation.

Regarding to computational efficiency, the application has been implemented in C++, using the OpenCV[12] library for some image processing operations. Tests have been executed on a general purpose PC (PIV, 3.0 GHz, 1GB-RAM), which is able to simultaneously run the application over sequences from three cameras with a resolution of 640x480 and 15 fps. The results can be checked online in the Medusa project web page: http://dymas.ii.uam.es/~agm/.

Compared to techniques just based on running gaussian average[7][8], foreground objects detected by our combined method have their borders more defined. So, misclassified holes inside these objects can be detected as real holes and included into the foreground. The combined method has also shown greater robustness against image noise.

In order to comparatively evaluate overall performance of the proposed approach, two other state-of-the-art methods have also been tested into the aforementioned application. The first one uses an improved GMM[3], and the second a statistical approach[11] (as implemented in the OpenCV library). Our results both for live video and for specifically generated test sequences, show that the proposed method outperforms them in several aspects: quality of segmentation at low frame rates, initialization time and

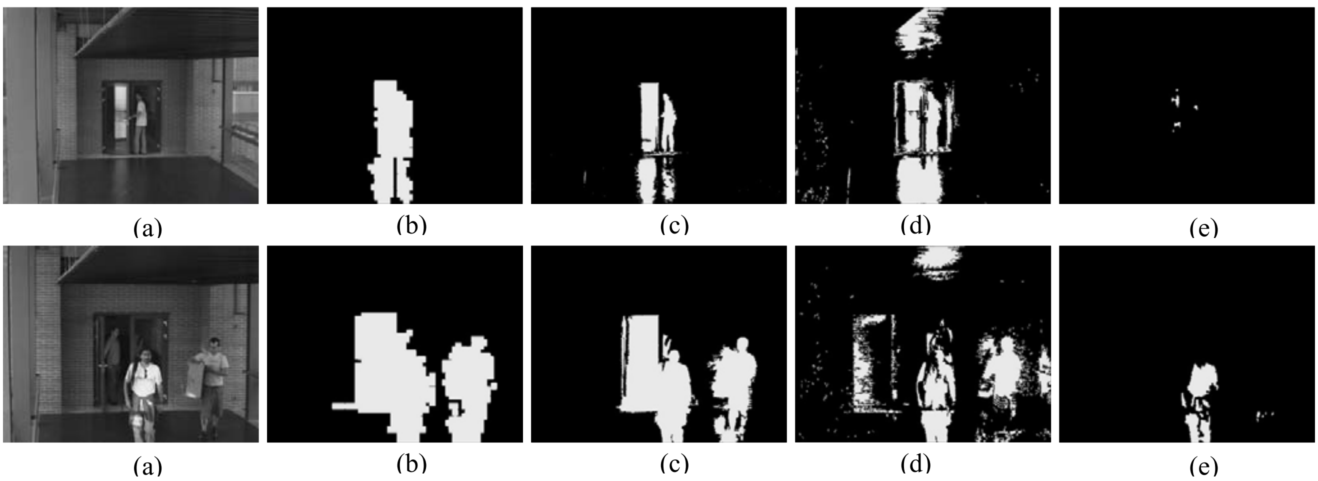

Figure 2: Segmentation results obtained for two frames in one of the test sequences.: (a) Current frame; (b) proposed method, 8x8 blocks; (c) proposed method, pixels; (d) improved GMM algorithm [3]; (e) statistical approach [11].

efficiency. Our algorithm has proven to obtain precise segmentation masks for high and low frame rates without modifying the value of its parameters (however, significant changes in the noise level require the adjustment of some of them). Figure 2 shows qualitative results of the achieved segmentation, while Table 1 shows measured efficiency for each tested method.

Image size	352x288	640x480
Proposed method (pixels)	62 fps	20 fps
Proposed method (8x8 blocks)	>300 fps	95 fps
GMM[3]	25 fps	8 fps
Statistical approach[11]	20 fps	5 fps

Table 1: *Efficiency for tested algorithms.*

Typical values for execution parameters are shown in Table 2: T_{tc}, is the threshold for frame difference, τ_{acc} the length of the adaptation window for background subtraction operations[8], k_0 the value for the confidence factor in the background subtraction described in section 2.1, and k_{max} and k_{min} are explained in section 2.2.

Symbol	Value	Meaning
T_{tc}	15-30	Temporal change threshold
τ_{acc}	100-200	Length of bkg adaptation window
k_0	2.5	Blind confidence factor
k_{max}	3-6	Non-changed point confidence factor
k_{min}	1-2	Changed point confidence factor

Table 2: *Parameter values for a typical indoor application.*

4. CONCLUSIONS

This paper proposes a method for foreground detection in video sequences, using temporal change detection to improve the results of a background subtraction method. Assuming a non-complex background, common in many indoor surveillance applications, the method is able to work in real-time while adapting to smooth illumination changes.

Results show a good compromise between efficiency, accurate segmentation and robustness, being able to work with high and low frame rates while complying with real-time requirements.

This method introduces a context-aware background subtraction method as its main innovation. In the current implementation, temporal change detection represents this context information. However, as pointed out, the framework allows further use of additional context information to improve the background subtraction with very little effort.

REFERENCES

[1] C. Stauffer, W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," *CVPR 1999*, p. 2246, 1999.

[2] P.W. Power, J.A. Schoonees, "Understanding Background Mixture Models for Foreground Segmentation", *Proceedings Image and Vision Computing New Zealand, 2002.*

[3] P. KaewTraKulPong, R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection", *Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, AVBS01. Sept 2001.*

[4] J. Rymel, J. Renno, D. Greenhill, J. Orwell, G.A. Jones, "Adaptive eigen-backgrounds for object detection," *Image Processing, 2004. ICIP '04. 2004 International Conference on* , vol.3, no.pp. 1847- 1850 Vol. 3, 24-27 Oct. 2004

[5] M. Piccardi, T. Jan, "Mean-shift background image modelling," *Image Processing, 2004. ICIP '04. 2004 International Conference on* , vol.5, no.pp. 3399- 3402 Vol. 5, 24-27 Oct. 2004

[6] B. Han, D. Comaniciu, L. Davis, "Sequential kernel density approximation through mode propagation: applications to background modeling", *Proc. ACCV – Asian Conf. on Computer Vision, 2004.*

[7] M. Piccardi, "Background subtraction techniques: a review," *Systems, Man and Cybernetics, 2004 IEEE International Conference on* , vol.4, no.pp. 3099- 3104 vol.4, 10-13 Oct. 2004.

[8] S. Huwer, H. Niemann, "Adaptive Change Detection for Real-Time Surveillance Applications" *Visual Surveillance, 2000. Proceedings. Third IEEE International Workshop on*, vol., no.pp.37-46, 2000.

[9] V. Mezaris, I. Kompatsiaris, N. Boulgouris, M. Strintzis, "Real-Time Compressed-Domain Spatiotemporal Segmentation and Ontologies for Video Indexing and Retrieval", *IEEE Trans. Circuits Syst. Video Technol.*, May 2004

[10] S.M. Desa, Q.A. Salih, "Image Subtraction for Real Time Moving Object Extraction," *Computer Graphics, Imaging and Visualization, 2004. CGIV 2004. Proceedings. International Conference on* , vol., no.pp. 41- 45, 26-29 July 2004.

[11] L. Li, W. Huang, I.Y.H. Gu, Q. Tia, "Foreground object detection from videos containing complex background", *Proceedings of the eleventh ACM international conference on Multimedia*, November 02-08, 2003, Berkeley, CA, USA.

[12] OpenCV, open source library for computer vision. *http://www.intel.com/technology/computing/opencv/overview.htm*