The Sakbot System for Moving Object Detection and Tracking

R. Cucchiara[1], C. Grana[1], G. Neri[2], M. Piccardi[3], A. Prati[1]

[1] D.S.I. - University of Modena and Reggio Emilia, via Vignolese 905 - 41100 Modena, Italy
[2] D.E.I.S. - University of Bologna, viale Risorgimento 2 - 40136 Bologna, Italy
[3] Dip. Ingegneria - University of Ferrara, via Saragat 1 - 44100 Ferrara, Italy

Abstract. *This paper presents Sakbot, a system for moving object detection and tracking in traffic monitoring and video surveillance applications. The system is endowed with robust and efficient detection techniques, which main features are the statistical and knowledge-based background update and the use of HSV color information for shadow suppression. Tracking is performed by means of a flexible tracking module based on symbolic reasoning, which can be tuned to several different applications.*

1. INTRODUCTION

In the last decade, many approaches have been proposed for moving object detection and tracking from videos, mainly dedicated to traffic monitoring and visual surveillance. Detection of moving objects in videos (moving visual objects, or MOVs for short hereafter) is based on the realistic assumption that in almost each frame MVOs are perceivable as different from background and from recent previous frames. In addition, models of target MVOs can be used for improving detection in those cases where target objects allow us a model-based description and models are known a priori. Instead, model-based detection cannot be performed in cases where precise geometrical models are not possible (for instance, for nonrigid objects like pedestrians) or, simply, object models are not exhaustively known in advance.

Tracking aims to describe trajectories of moving objects during time. The main problem to solve for tracking is to find correspondences of the same physical objects in different frames. This problem may result either trivial or hard depending on several aspects of the scene, such as the density and proximity of objects, rigid or variable shapes, presence of occlusions, and others.

In this work we aim to detect and track different and possibly unknown objects, including vehicles and pedestrians in outdoor traffic scenes but also casual walkers in parking or private areas. The approach we propose (Sakbot, Statistic and Knowledge-based Object Tracker) is meant to be general-purpose and thus adequate for application ranging from vehicle tracking to visual surveillance. Unlike other works in the field, we do not address the relatively simpler case of road surveillance, where one class of moving objects (i.e., vehicles) can be detected along some fixed motion directions. We consider instead more complex situations with moving people and vehicles, with different shape, speed, trajectory, presence of infrastructures and possible occlusions.

Sakbot is conceived as a general-purpose system adopting a single fixed camera, able to deal with several different operational situations:

a) luminance condition changes, due to atmospheric conditions, day hours, shadows.

b) limited, high-frequency camera motions, due to vibrations and the wind, which are compensated in a pre-processing phase;

c) background changes; the background of the scene could change either because objects are stopped (e.g. a stopped vehicle) or because still objects (previously considered background objects) start their motion.

d) MVOs can move with whatever trajectory and speed; thus our approach cannot be based on frame difference, a method that should calibrate difference in dependency on object speed. At the same time, motion models cannot be exploited.

In this context the process must be featured with a general and robust MVO detection phase and a flexible tracking and reasoning system capable of adapting to different applications. For the detection phase, we use a background suppression technique with selective adaptation of the background, improved by the knowledge of motion at object level rather then at pixel level. For the tracking phase, we exploit a framework of symbolic rules governing the coherence of detected MVOs in the scene and during time.

The rest of the paper is organized as follows: Section 2 presents the main recent related works. Section 3 describes the Sakbot system with sub-sections for the architecture, the background estimation, and the shadow suppression techniques. Section 4 describes the Sakbot implementation and main application. Conclusive remarks addresses the main results and applications of Sakbot.

2. RELATED WORKS

Many systems have been proposed recently in the literature for moving object detection and tracking. These systems obviously differ in the approach proposed, but also in various assumptions about the operational environment. One first main distinction is between systems adopting a single, fixed camera [1-6] with respect to systems adopting either multiple cameras [7] or an airborne camera [8]. In this work, we focus on a single fixed camera scenario, since it still captures a wide spread of applications. Another relevant distinction is made between systems oriented to monitoring of traffic scenes, where the main targets are vehicles and pedestrians (see for instance [1,5,6]), and systems for video surveillance of unattended areas such as metro platforms or parking areas (see [3, 4]). In the two cases, different a-priori knowledge about objects in the scene can be exploited in order to improve detection and/or tracking. In this work, we assume that both vehicles and pedestrians must be detected and tracked, adopting a flexible symbolic reasoning module for the tracking phase. The reasoning module is based on a set of rules which allows tracking

of moving objects during time, granting coherence of object trajectories exploiting the semantic of the scene. This feature is similar to the use of predefined scenarios and a-priori contextual information described in [3]. In our work, the set of rules was initially tuned to vehicle tracking [6], but was then easily augmented to consider pedestrians in traffic scenes and casual intruders of predefined areas.

3. THE SAKBOT SYSTEM

The main problem to be solved for the moving object detection phase is the definition of a fast and robust background update technique. To this aim, we have defined a specific approach, called S&KB background update, based on a statistical adaptation of the background, together with a "knowledge-based" selective background update [9]. The statistical adaptation copes with the changing appearance of the scene during time, while the knowledge-based selection prevents updating the background with foreground points. Thank to adaptation, the system is able to integrate long stationary MVOs (e.g. a parked vehicles) or areas abandoned by previously still objects into the background; selectivity allows the system to discriminate apparently moving objects from real ones.
The main features of Sakbot are:

- the use of *color information* (instead of grey levels only), during the entire MVO detection process, which improves sensitivity;
- the use of a *median function with adaptivity* that is a good approximation of a mode function and assures high responsiveness to background changes and good accuracy, even if few time-samples are considered;
- the inclusion of a *knowledge-based* feedback from the segmentation module in order not to insert in the background parts of MVOs; this selective update, based on knowledge of whole objects and not of single pixels only, abates false positives (background pixels recognized as MVO pixels) and prevents from deadlock situations in background update;
- the exploitation of a *shadow detection module* for improving both object segmentation and background updating.
- The definition of a set of flexible rules working on the moving visual objects for providing tracking and classification task that can be adapted to different scenarios.

For a general-purpose video surveillance system, a flexible tracking and reasoning module capable of adapting to different applications is needed. Therefore, we developed the tracking level with a framework of symbolic rules which can be easily understood and updated by human experts.

3.1. The Sakbot architecture

The architecture of Sakbot consists of several steps (see Fig. 1). The processing time for these steps determines the upper-bound of the achievable processing rate.

Each new input frame (IF_t) undergoes a camera motion correction step, which uses a calibrated fixed point in the frame and a standard frame-by-frame cross-correlation for adjusting limited but unavoidable camera movements. The corrected image I_t is computed by shifting the correspondent sampled frame IF_t so that the fixed point is juxtaposed with the minimum of the correlation function.

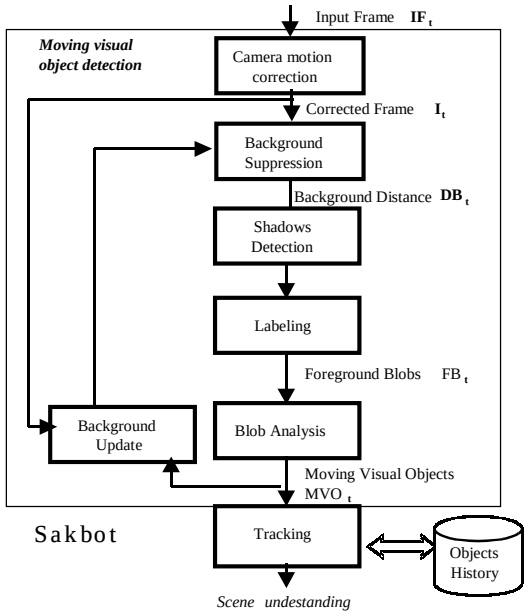


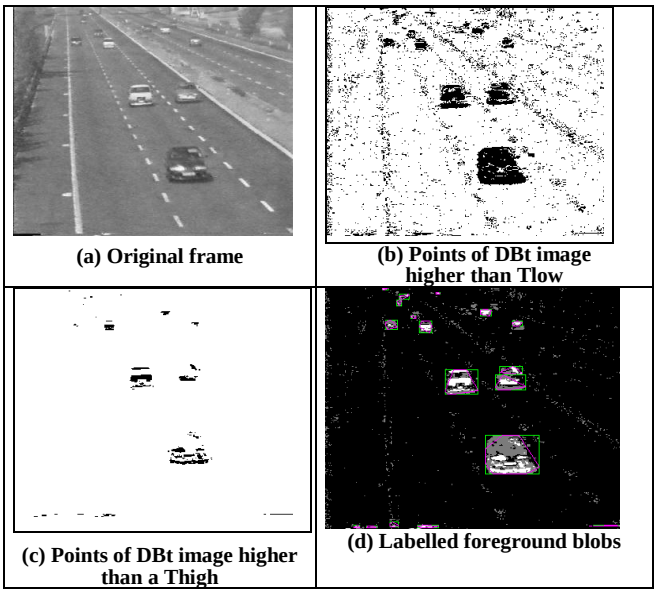
Fig. 1. The Sakbot architecture.

Once the input image has been corrected, the still background is subtracted to detect MVOs. Differently from many other proposals (see for instance [1,6]), we exploit the information given by the pixel chrominance to better discriminate foreground from background.

The difference image DB_t (Difference with Background) is computed as follows:

$$DB_t(x,y) = Distance(I_t(x,y), B_t(x,y)) = \max(|I_t(x,y).c - B_t(x,y).c| \; c=R,G,B) \quad (1)$$

Therefore the distance image results to be a scalar, grey level-like image containing only candidate moving points.



(a) Original frame

(b) Points of DBt image higher than Tlow

(c) Points of DBt image higher than a Thigh

(d) Labelled foreground blobs

Fig. 2. Threshold with hysteresis.

Many background suppression approaches obtain the moving points by simply thresholding the DB_t image. In order to limit the dependency of results from a given threshold, we adopt a threshold with hysteresis, conceptually similar to that exploited in [10]. In [10] points are selected if greater than a low threshold and belonging to a neighbourhood of points higher than a high threshold. In this work, we use threshold with hysteresis added with morphology: we first select moving points with a low threshold Tlow (Fig. 2.b); then, morphological operators (closure and opening) are applied on the image in order to eliminate isolated points or very small spots due to noise. Then labelling is performed, by accepting blobs containing at least one point greater than a high threshold Thigh (see Fig. 2.c). The labelled image at time t thus contains a number of Foreground Blobs (FBt image, Fig. 2.d).

On the segmented foreground blobs we perform blob analysis, consisting of two steps:

1) blobs with an area less than a TAREA threshold (which depends on the distance between camera and scene and on the typical size of objects) are discarded;

2) the average optical flow (aOF) is computed for each blob. All blobs with aOF less than a TOF threshold are discarded. They are considered as apparently moving objects (due to a locally wrong background) and thus not accepted as real MVOs.

Finally, a set of geometrical features is computed on the detected MVOs in order to use this information in the tracking module.

3.2. The Sakbot background update

The model of background we defined is based on the statistical assumption that the background points should be the most probable points observed in a finite window of observation time. As statistical function we use the median of sampled frames, that we have proven to be effective as the mode function but requiring a limited number of frames, and thus being less time and memory consuming. In the literature, the background is usually computed with an adaptive function which takes into account past background values and the current frame. Accordingly, in order to improve the stability of the background update, we include an adaptive factor, i.e., the previous background weighted with an adequate weight wb. Therefore we define:

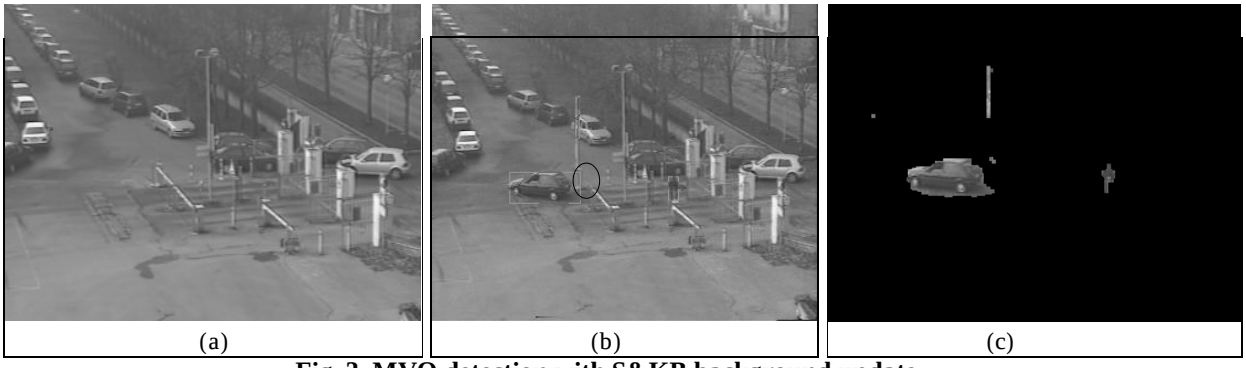
$$Bt = median\ (It-\Delta t,\ldots,It-n\Delta t\ ,wbBt-\Delta) \quad (2)$$

With the use of statistics for background computation, MVO points should not be included in the background, thank to their low occurrence. Nevertheless, many errors occur, especially when objects are large, homogeneous in color, and slow: in this case, points of MVOs could turn out to be included in the statistical.

The statistical and adaptive update could be improved by a selective update, which does not update the background value if the point is marked as moving in the current frame, as in [11]. The drawback is the risk of a deadlock: if for any reason a fixed point is incorrectly detected as moving, it will remain excluded from background update, and therefore detected as moving point forever. In order to avoid this problem we exclude points belonging to detected MVOs after a further validation: MVOs must have a non null average optical flow aOF. Knowledge about the *whole* moving object makes selection reliable. Thus we define the S&KB background update for each point in Bt as

$$Bt = B(t-1)\ if\ It \in MVOk\ \&\ aOFk > TOF\ for\ any\ k,$$
$$= median\ (It-\Delta t,\ldots,It-n\Delta t\ ,wbBt-\Delta)\ \ otherwise \quad (3)$$

Fig. 3 shows the effect of adding the feedback based on the knowledge of MVOs. Fig. 3.a represents the background with three closed barriers. In Fig. 3.b the bar has risen and a car is passing. In this frame four MVOs are detected: the car, the people, the rising bar and a virtual bar in the horizontal position (in the black circle). This object is a false positive and can be discriminated since its OF is null. Therefore points in the area occupied by the virtual bar will be updated in the computed background with actual background values. Instead, points of the other three MVOs will be masked (and validated as real MVOs, see Fig. 3.c).



(a) (b) (c)

Fig. 3. MVO detection with S&KB background update.

3.3. Shadow detection in Sakbot

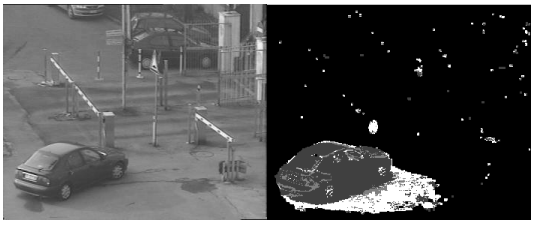
A shadow detection module has been added for improving both MVO detection and background update. Shadow detection is performed based on chrominance analysis, applied to points belonging to moving objects only, in order both to limit the computation and discard fixed shadows belonging to the background.

To detect shadow, we first convert from the color space (R,G,B) to (H,S,V). The color space (H,S,V) better reproduces the human visual behavior and it is more sensitive to brightness changes due to shadows. By exploring many videos with different light conditions we observed that a pixel “covered” by shadows

becomes darker (V component), but also exhibits a color saturation (H and S components). Thus, for a shadowed point (x,y) we have:

$$SP_t(x,y)=\begin{cases}1 & \text{if } \alpha \le \dfrac{I_t(x,y).V}{B_t(x,y).V} \le \beta \ \wedge\ |I_t(x,y).H - B_t(x,y).H| \le \tau_H \\ & \wedge\ |I_t(x,y).S - B_t(x,y).S| \le \tau_S \ \wedge\ FPm_t^{TL}=1 \\ 0 & \text{otherwise}\end{cases}$$

with $0 < \alpha,\beta,\tau_H,\tau_S < 1$. Intuitively, this means that shadow darkens a "covered" point, but *not too much*. Typically β ranges from 0.9 to 0.97; higher β values cause shadow detection to be too sensitive to noise; instead, α ranges from 0.75 to 0.85, meaning that typically a "shadowed" point becomes darker of about 20%. Typical values for τ_H and τ_S are 0.15. The condition on the V component is similar to the shadow detection technique proposed in [11].



(a) (b)

Fig. 4. Segmentation with shadow detection.

Fig. 4 shows an example of the improvement achievable with the method we use in Sakbot: Fig 4.a shows the original frame; in Fig. 4.b, background pixels are colored in black, foreground pixels in dark grey, shadow pixels detected if using only the condition on the V component in light grey, and eventually shadow pixels detected adding the H and S conditions, in white. In Fig. 4, it is possible to note that the car is partially eroded by light grey pixel, i.e. using color information reduces false positives in shadow detection; moreover, real shadows are detected more effectively.

4. SAKBOT IMPLEMENTATION AND APPLICATION

The current Sakbot prototype has been developed using the Microsoft Visual C++ 4.0 environment and ANSI C libraries as an executable for standard PCs with MS Windows. In the current version, acquisition and processing are performed separately. The input video format supported are AVI and MPEG-2. The average frame rate is ten frames per second with reduced (360 x 288) frame size, of course depending directly on the PC's performance.

Test videos have been acquired with a 3-CCD PAL-DV digital camera (Sony DCR-VX9000E), which can be connected to the PC via a standard IEEE 1394 connection. The PAL-DV raw format (720 x 576) does not provide square pixels, but the ratio between height and width is limited (the pixel horizontal size is (4/3)/(5/4) = 1.064 larger than vertical size), thus pixels were assumed square in processing.

Background is updated with a ΔT of 10 frames, so that a new background is computed every 1 second. All the other operations, including background subtraction, segmentation, shadow suppression and motion validation via optical flow are computed at frame rate.

The first Sakbot prototype has been validated in the Campus of the University of Modena, Italy, for people and vehicle surveillance at an access gate and will be used as an experimental system for urban traffic control, installed on traffic-light controllers in Bologna City. In this context, the research has been funded by the "Supervisor of traffic control of Bologna" project co-financed by the Bologna Municipality and Italian Government.

5. CONCLUSIONS

This paper has presented Sakbot, a general-purpose system for detection and tracking of moving objects in traffic monitoring and video surveillance applications. Its architecture is based on an effective and efficient detection module and a flexible tracking module based on symbolic reasoning. In this paper, a description is given of the background update and shadow suppression algorithms. Sakbot is currently experimented in the context of a traffic monitoring application at the University of California, San Diego, and will be used in a project for sustainable mobility of the Town and Province administrations of Bologna, Italy, and the local public transportation company (ATC).

REFERENCES

[1] Lipton, A., Fujiyoshi, H., Patil, R., "Moving target classification and tracking from real-time video", Proc. of WACV '98, pp. 8-14, 1998.

[2] N. Chleq, M. Thonnat, "Realtime image sequence interpretation for video-surveillance applications", Proc. of ICIP96, pp. 59-68, 1996.

[3] N. Rota, M. Thonnat, "Video sequence interpretation for visual surveillance", Proc. of Third IEEE Int. Workshop on Visual Surveillance 2000, 2000, pp. 59-68.

[4] F. Brémond, M. Thonnat, "Tracking multiple nonrigid objects in video sequences", IEEE Trans. on Circ. and Syst. for Video Tech., v. 8, n. 5, 1998, pp. 585-591.

[5] G. L. Foresti, "Object recognition and tracking for remote video surveillance", IEEE Trans. on Circ. and Syst. for Video Tech., v. 9, n. 7, 1999, pp. 1045-1062.

[6] R. Cucchiara, M. Piccardi, P. Mello, "Image analysis and rule-based reasoning for a traffic monitoring system", IEEE Trans. on Intelligent Transportation Systems, Vol. 1, No. 2, June 2000, pp.119-130.

[7] J. Orwell, P. Remagnino and G.A. Jones "Multi-camera color tracking", Proc. of the 2nd IEEE Workshop on Visual Surveillance, pp. 14-22, 1998.

[8] I. Cohen, G. Medioni, "Detecting and tracking moving objects for video surveillance", Proc. of CVPR 99, 1999, pp. 319-325.

[9] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, "Statistic and knowledge-based moving object detection in traffic scenes", in Proc. of ITSC-2000 - 3rd IEEE Conference on Intelligent Transportation Systems, pp. 27-32, 2000.

[10] M. P., Dubuisson Jolly, S. Lakshmanan, A. K. Jain, "Vehicle segmentation and classification using deformable templates". IEEE Trans. Pattern Anal. Machine Intell., vol. 18, no. 3, 1996, pp. 293-308.

[11] A. Elgammal, D. Harwood, L. Davis, "Non-parametric model for background subtraction," in Proc. of FRAME-RATE, Corfu, Greece, 1999.