UNSUPERVISED STATISTICAL DETECTION OF CHANGING OBJECTS IN CAMERA-IN-MOTION VIDEO

Rozenn Dahyot, Pierre Charbonnier

Laboratoire Régional des Ponts et Chaussées
11 rue Jean Mentelin
67035 Strasbourg, France

Fabrice Heitz

LSIIT CNRS / Strasbourg University
Boulevard Sébastien Brandt
67400 Illkirch, France

ABSTRACT

Change detection in image sequences has mainly focused on the recovery of moving objects when the viewing system is static [1], or on the detection of simple production effects such as video shot boundaries or scene transitions [2]. Camera motion is usually handled by the compensation of dominant motion, using motion estimation and segmentation schemes [1]. In this paper we propose a novel statistical change detection method able to handle more complex events such as entering or exiting objects, or changes in objects appearance, when the camera is moving. Temporal changes of objects are captured by analyzing the statistics of successive images. Considering an appropriate choice of image features, we show how it is possible to extract the statistics of changing objects from a pair of successive image histograms. Changing objects are then located by statistical back-projection techniques. The method is completely unsupervised and does not require any motion estimation or motion compensation. It is illustrated here on real world road scenes exhibiting large camera motion.

1. INTRODUCTION

Automatic video analysis is an essential task for many applications, such as content-based video retrieval or content-based video coding. In this paper, we propose a new statistical change detection method, able to handle complex temporal changes in camera-in-motion video. Statistical approaches have already shown efficient for scene transition detection [2, 3], yielding fast, unsupervised methods, and providing insensitivity to camera motion. Here, we extend this concept to the detection of the more complex events involved by exiting, entering or changing objects in one-shot sequences.

We assume that each frame of the shot is composed of a set of fast changing objects O and of a slowly moving background $\overline{O}$. In practice, we are interested in changing objects that appear or disappear in the video, or objects which exhibit local features that change significantly over time. Under appropriate assumptions on O and $\overline{O}$, we show in section 2.2 how to extract the probability density function (p.d.f) of changing objects. Statistical back-projection techniques are then used to locate the objects of interest, characterized by their p.d.f. [4, 5] (section 3).

The detection procedure is unsupervised and fast since it does not require image matching or motion correspondence. The method is applied to a multi-template detection problem in camera-in-motion video sequences of road scenes (section 4).

2. STATISTICAL CHANGE DETECTION

We consider a sequence of images $I(t)$ of a scene which is composed of two classes: the background $\overline{O}$, which represents its major part, and the objects of interest O. While the statistical variations of the background over time are assumed slow, the objects of interest strongly change in size or in appearance. They may even enter or exit the scene. In contrast to supervised methods [6, 7], we do not perform any *a priori* learning of the statistics of the object class. However, we assume that we can define a local measurement m, which allows to discriminate objects from the background. In this section, we first show how to put these assumptions into mathematics. Then, we show how it is possible to extract the statistics of changing objects $f(m,t|O)$ from the statistics of two successive images $f(m,t)$ and $f(m,t+1)$.

2.1. Statistical assumptions on the scene

As the scene is composed of two classes, namely the background and the objects, it is natural to write the distribution $f(m,t)$ of a local image measurement m at time t as a mixture of two conditional p.d.fs:

$$H_0: \quad f(m,t) = (1-\epsilon(t))f(m,t|\overline{O}) + \epsilon(t)\, f(m,t|O) \quad (1)$$

where $\epsilon(t)$ is the proportion of the scene belonging to the object.

• The first assumption H_1 is that the background is the dominant region in the scene: $\epsilon(t) < 0.5 \; \forall t$.

• The second assumption is that the background statistics change slowly over time:

$$H_2: \quad f(m,t|\overline{O}) \simeq f(m,t+1|\overline{O}) \tag{2}$$

• The third assumption is that the chosen image measurement, m, captures the difference between objects and background. Therefore:

$$H_3: \begin{cases} if\ f(m,t|\overline{O}) > 0\ then\ f(m,t|O) \simeq 0 \\ if\ f(m,t|O) > 0\ then\ f(m,t|\overline{O}) \simeq 0 \end{cases} \tag{3}$$

• Let us recall that our goal is to detect changing objects, i.e. objects such that $\epsilon(t)\ f(m,t|O)$ significantly changes between two successive images $I(t)$ and $I(t+1)$. Two cases are considered:

The first case corresponds to significant changes in object appearance. A significant change is defined if the following conditions are verified:

$$H_4: \begin{cases} if\ f(m,t|O) > 0\ \ then\ \ f(m,t+1|O) \simeq 0 \\ if\ f(m,t+1|O) > 0\ \ then\ \ f(m,t|O) \simeq 0 \end{cases} \tag{4}$$

The second case of interest corresponds to significant changes in object size i.e. $H_4^{'}: \epsilon(t)$ "changes", while $f(m,t+1|O) \simeq f(m,t|O)$. This occurs for example when objects enter or leave the scene. The definition of "significant changes" in this second case is made clear in the next section.

2.2. Distribution of changing objects

2.2.1. Extraction of $f(m,t|O)$

Let us consider function, $\Delta_+(m,t)$ defined by:

$$\Delta_+(m,t) = \max\left(f(m,t) - 2\ f(m,t+1), 0\right) \tag{5}$$

defined as the "backward extractor". Using H_0, H_1, H_2 and H_3, we show (see appendix A for details) that *if H_4 is satisfied:*

$$\Delta_+(m,t) = \epsilon(t)\ f(m,t|O), \tag{6}$$

and *if $H_4^{'}$ is satisfied:*

$$\begin{cases} \Delta_+(m,t) = 0 \quad if\ \epsilon(t) < \ 2\ \epsilon(t+1) \\ \Delta_+(m,t) = (\epsilon(t) - 2\epsilon(t+1))\ f(m,t|O) \ \ otherwise \end{cases} \tag{7}$$

As a consequence, since $f(m,t|O)$ integrates to one, it can directly be derived from Δ_+:

• if the appearance of O changes (H_4),

• or if the proportion $\epsilon(t)$ of the objects in the scene *decreases* by a factor of two or more between $I(t)$ and $I(t+1)$ ($H_4^{'}$).

Similarly, we can define a "forward extractor" by:

$$\Delta_-(m,t) = \max(f(m,t) - 2\ f(m,t-1), 0) \tag{8}$$

to detect events between $I(t)$ and its previous image $I(t-1)$ such as:

• objects with changing appearance,

• objects the proportion of which *increases* by a factor two or more.

2.2.2. Remarks

Let us notice that the objects detected by the backward and forward extractor Δ_+ and Δ_- are not necessarily the same. Δ_+ detects *changing* objects in $I(t)$ by comparison with $I(t+1)$ and Δ_- detects *changing* objects in $I(t)$ by comparison with $I(t-1)$. Δ_+ is well suited to detect *disappearing* objects *from* $I(t)$ while Δ_- is suited for the detection of structures *appearing in* $I(t)$. Thus, the two detectors are complementary.

3. LOCALIZATION OF STATISTICAL CHANGES

In order to locate changing objects in $I(t)$, we compute a confidence map. At every pixel (i,j), we extract a local measurement m_{ij}, and estimate the *a posteriori* probability that the pixel belongs to a changing object, according to the local measurement value, i.e.:

$$f\left(O|m_{ij},t\right) = \frac{f\left(m_{ij},t|O\right)\ f\left(O\right)}{f\left(m_{ij},t\right)} \tag{9}$$

where $f(O)$ the prior p.d.f. of objects is assumed to be constant, $f(m,t)$ is approximated by the histogram of measurement m in $I(t)$ and $f(m,t|O)$ is obtained as described in section 2.2. This probability is stored into a confidence map, at position (i,j). This approach has been called the "back-projection" technique in object detection in static images [4, 6]. We call our method backward back-projection when Δ_+ is used and forward back-projection when Δ_- is used.

The localization procedure is summarized in fig. 1: p.d.fs of three successive images are computed as multidimensional histograms. Then, Δ_+ and Δ_- are computed and both backward and forward back-projections are performed.

4. EXPERIMENTAL RESULTS

In this experiment, we consider large video databases collected on the French road network for infrastructure safety studies. In this context, frames are not indexed by time, but by curvilinear abscissa (one image every 5 meters): the camera motion is large, and corresponds mainly to forward travelling. The target class O comprises still objects located

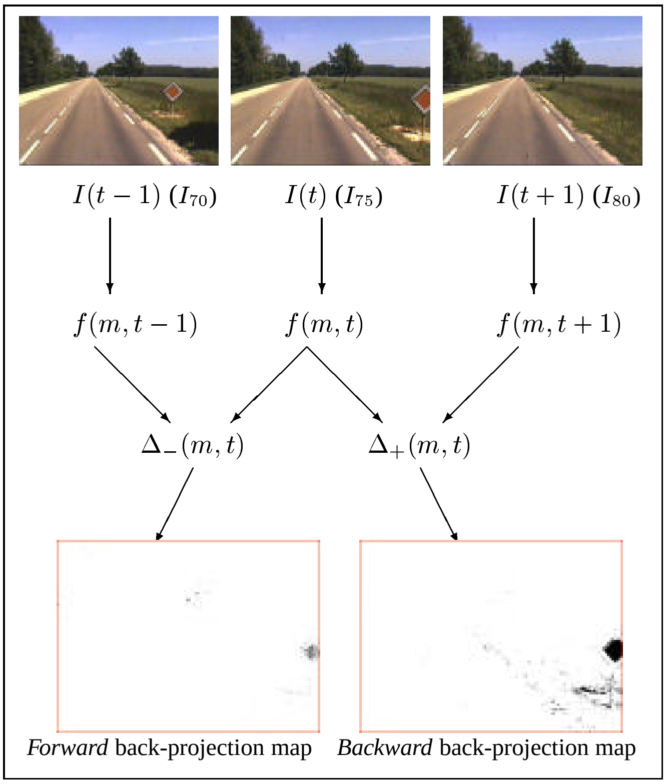


Fig. 1. **Detection of changing objects:** changing features appear as dark pixels in the back-projection maps of I_{75}. Color features have been used here (see text).

on the roadside, such as trees, poles, delineators and road signs. In order to satisfy assumption H_3, two kinds of measurements, m, have been selected for this multi-template detection problem:

• the first one relies on color, which is a useful attribute in the case of road signs [8],

• the second one is related to shape, since many manufactured objects, but also tree trunks, have straight contours.

We apply in both case backward and forward back-projection, as a quick detection procedure, on 200 successive 768×564 color images. Typically, the computation time is about one second per image.

4.1. Color features

The chosen color features are chromatic coefficients (r, g) extracted from RGB pixel values:

$$m_{color} = \begin{pmatrix} r = \frac{R}{R+G+B} \\ g = \frac{G}{R+G+B} \end{pmatrix} \tag{10}$$

Fig. 1 shows the confidence maps obtained for image I_{75}, with the color features: white pixels correspond to zero probability while dark pixels correspond to high confidence. Though no thresholding is applied here, the proportion of zeroes in the maps is large, showing that the method is discriminant. The yellow road sign grows between images I_{70} and I_{75}. Its proportion increases more than twice, while its color does not change. We are in the $H_4^{'}$ case and the road sign is detected in the forward back-projection map. The road sign disappears between images I_{75} and I_{80}: $H_4^{'}$ is, again, satisfied and the object is detected by the backward procedure.

4.2. Shape features

Three parameters have been used to define a shape measurement, related to the alignment of local edges. The first parameter is the angle θ of a local edge. The second one called α, is an alignment measurement (two points belonging to the same straight contour have identical (θ, α) values). The third parameter is the norm N of the gradient:

$$m_{shape} = \begin{pmatrix} \theta = \arg\tan \frac{I_x}{I_y} \\ \alpha = x\,\frac{I_x}{N} + y\,\frac{I_y}{N} \\ N = \sqrt{I_x^2 + I_y^2} \end{pmatrix} \tag{11}$$

where I_x and I_y denote the two components of the spatial gradient. Figure 2 shows the forward and backward back-projection maps of I_{75} for shape measurements. Disappearing road sign contours are well detected in the backward back-projection (assumption $H_4^{'}$). Since the value α is sensitive to the position (x, y) in the image, the forward back-projection map also discriminates road sign contours because their shape features change, so assumption H_4 is satisfied. Let us notice that the method is different from edge detection: only the contours of changing objects are detected. As an illustration, fig. 3 shows image I_{405} along with its contours and back-projection maps. While the contours of the delineator are outlined in the back-projection maps, those of road markings - that remain statistically the same throughout the sequence - are not.

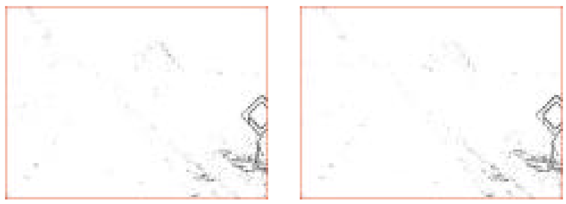
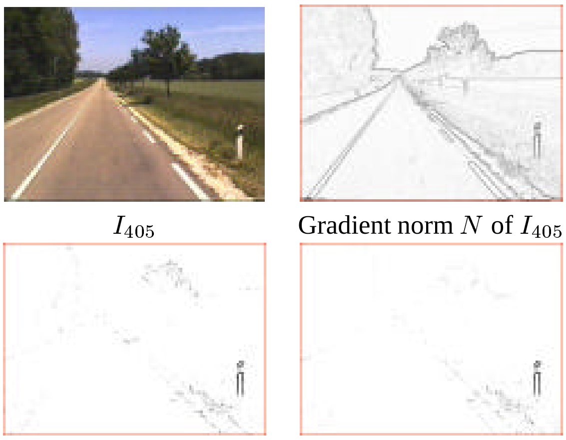


Fig. 2. Forward and backward back-projection maps of I_{75} with shape measurements

4.3. Experimental results

In the - 200 frames - test sequence considered here, there are 5 road signs (four of them have colored parts, one is black and white) and 19 delineators (see example fig. 3). We consider an object as detected if most pixels of either its colored part or its straight contours have non-null probability in one of its back-projection maps, depending on the considered measurement. As expected, the four colored road signs were detected using color features (see fig. 1) and all five road signs were detected using shape measurements (see fig. 2). Color features are not suited for delineator detection ($0/19$ found) but shape features are ($15/19$ found). The four missing detection are explained by strong shadows that completely hide the contours of the delineators.



I_{405} — Gradient norm N of I_{405}

Forward map of I_{405} — Backward map of I_{405}

Fig. 3. Detection results on I_{405} with shape measurements

5. CONCLUSION

We have shown that relevant statistical information about changing objects can be extracted from image sequences, with moving camera. The associated object detection method is fast, fully unsupervised and may be adapted to the detection of different objects by selecting appropriate local image measurements.

6. REFERENCES

[1] A. Mitiche and P. Bouthemy, "Computation and analysis of image motion: A synopsis of current problems and methods," *Int. Jl. Computer Vision*, vol. 19, no. 1, pp. 29–55, 1996.

[2] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classifying production effects," *Multimedia systems*, vol. 7, pp. 119–128, 1999.

[3] J.S. Boreczky and L.A. Rowe, "Comparison of video shot boundary detection techniques," *In Storage and Retrieval for Image and Video Databases IV, proc. SPIE 2670*, pp. 170–179, 1996.

[4] J.R. Smith, *Integrated Spatial and Feature Image Systems: Retrieval Analysis and Compression*, Ph.D. thesis, Columbia University, USA, 1997.

[5] M.J. Swain and D.H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[6] B. Schiele and A. Pentland, "Probabilistic object recognition and localization," in *proc. of International Conference of Computer Vision ICCV'99*, 1999.

[7] H. Schneiderman and T. Kanade, "A statistical method for object detection applied to faces and cars," in *proc. of conference of Computer Vision and Pattern Recognition CVPR'00*, Hilton Head, USA, June 2000, vol. 1.

[8] D.S. Kang, N.C. Griswold, and N. Kehtarnavaz, "An invariant traffic sign recognition system based on sequential color processing and geometrical transformation," in *IEEE International Conference on Image Processing ICIP'94*, Texas, 1994.

A. COMPUTATION OF Δ_+

A.1. Under H_4 assumption

Thanks to assumption H_0, we can write:

$$\begin{aligned} f(m,t) - 2\,f(m,t+1) &= \\ &(1-\epsilon(t))\,f(m,t|\overline{O}) - 2\,(1-\epsilon(t+1))\,f(m,t+1|\overline{O}) \\ &+\epsilon(t)\,f(m,t|O) - 2\,\epsilon(t+1)\,f(m,t+1|O) \end{aligned} \tag{12}$$

Since $f(m,t+1|\overline{O}) \approx f(m,t|\overline{O})$ thanks to H_2 assumption, equation (12) is simplified as:

$$\begin{aligned} f(m,t) - 2\,f(m,t+1) &= \\ &((1-\epsilon(t)) - 2\,(1-\epsilon(t+1)))\,f(m,t|\overline{O}) \\ &+\epsilon(t) f(m,t|O) - 2\,\epsilon(t+1)\,f(m,t+1|O) \end{aligned} \tag{13}$$

Assumptions H_3 and H_4 mean that if one of the three terms $f(m,t|\overline{O})$, $f(m,t|O)$ and $f(m,t+1|O)$ is not null, then the other ones are. Thus:

$$\begin{aligned} |f(m,t) - 2\,f(m,t+1)| &= \\ &|\,(1-\epsilon(t)) - 2\,(1-\epsilon(t+1))\,|\,f(m,t|\overline{O}) \\ &+\epsilon(t)\,f(m,t|O) + 2\,\epsilon(t+1)\,f(m,t+1|O) \end{aligned} \tag{14}$$

The last assumption H_1 yields:

$$|(-1 + 2\,\epsilon(t+1) - \epsilon(t))| = 1 - 2\,\epsilon(t+1) + \epsilon(t) \tag{15}$$

Since $\max(a,0) = \frac{a+|a|}{2}$, we have:

$$\Delta_+(m,t) = \epsilon(t)\,f(m,t|O) \tag{16}$$

A.2. Under $H_4^{'}$ assumption

In this case, we have:

$$\begin{aligned} |f(m,t) - 2\,f(m,t+1)| &= \\ &|\,(1-\epsilon(t)) - 2\,(1-\epsilon(t+1))\,|\,f(m,t|\overline{O}) \\ &+|\epsilon(t) - 2\,\epsilon(t+1)|\,f(m,t|O) \end{aligned} \tag{17}$$

Therefore if $\epsilon(t) > 2\epsilon(t+1)$, we have:

$$\Delta_+(m,t) = (\epsilon(t) - 2\epsilon(t+1)) f(m,t|O) \tag{18}$$

and $\Delta_+(m,t) = 0$ otherwise.