A Long Term Change Detection Method
for Surveillance Applications

Carlo S. Regazzoni, Andrea Teschioni and Elena Stringa
Department of Biophysical and Electronic Engineering
University of Genoa - Via Opera Pia 11A - 16145 Genoa - Italy
Phone +39-10-3532792 Fax +39-10-3532134
E-mail: andretes@dibe.unige.it

Abstract

A Long Term Change Detection (CD) Method is presented by definition of a probabilistic model and the integration of two different informative sources. The model is described from a theoretical point of view and its real implementation by means of a bank of shift registers is presented. The algorithm is part of a surveillance system for unattended railway stations; results on a real image sequence confirm its validity.

I. Introduction

The problem of detecting changes within an image sequence is often addressed in literature as an important one: in particular, the purpose of such kind of algorithms is the identification of areas having significative variations within an image sequence, providing then the a low level type information.

Change detection (CD) algorithms [1] consider as input two digitised images and provide as output a binary image where locations associated with differences are identified as non-null values: these differences may be caused by the motion of an object in the field of view, the addition or removal of an object from the observed scene, changes in illumination, or noise from the digitisation process. Robust change detection implies the capability of evaluating and discarding noise effects from the image of changes.

A preliminary classification of CD methods can be made between *short-term* change detection and *long-term* change detection; short-term algorithms make a decision whether a pixel should belong or not to a changed area only on the basis of the differential information with respect to a reference image acquired at close instant. Long-term methods analyse the information associated within an extended temporal window of an image sequence: the decision about the presence of a change is performed by considering frames within the sequence that show a temporal correlation with the current one.

CD methods receive as input two images and provide usually as an output a binary image where detected changes are represented as non-null values. Moreover, when one has to extract changes within a real-world image sequence, two different classes of change detection algorithms can be individuated, i.e. *frame-by-frame* [2] and *frame-background* [3] methods. In the first case the reference image is the previous frame of the sequence; in the latter case a background image is chosen as reference image, e.g. an image representing the scene whose changes must be detected.

The problem of detecting significant changes in dynamic real scenes has received more and more growing attention during last years and, actually, this problem arises in several areas of computer vision and signal processing.

For what concerns short-term CD algorithms, they have been commonly used for the detection of objects, both static both moving, within image sequences; such algorithms takes the advantage to be computationally quite simple, but they normally are less robust to noisy situations, as quickly illumination changes can be.

In order to avoid it, a certain quantity of information is tried to be integrated before that a decision about the presence or the absence of the object that has to be detected is taken.

In [4], for example, an adaptive algorithm has been developed for changes detection by recursively updating the grey-level value of each pixel, while [5] developed a multi-sensor algorithm which takes into account of the correlation coefficients between pairs of consecutive image frames.

In [6], moving people were detected by first estimating the background over a long image sequence and then by thresholding the difference between current frame and estimated background.

In this paper, a long-term spatial filter is described: the proposed method integrates two informative sources and two hierarchically organized steps are performed. A first step is done by applying a frame-by-frame algorithm and a frame-background algorithm in a parallel fashion. A further thresholding step is performed on the basis of the collected vectorial information.

This paper is organized as follows: section II and III show a general description of the proposed method and its probabilistic representation; in section IV the performances of the algorithm are presented: in particular, the optimization process of the *ta* threshold and some visual results are illustrated. Section V concludes this paper.

II. Detection of Abandoned Objects

The proposed long-term CD algorithm aims at identifying the presence and the location of abandoned objects inside a crowded environment (e.g. waiting room of a railway station). In particular, the proposed method points out static areas starting from a sequence of images areas which show relevant permanent differences with respect to a background scene (e.g. the empty room). The algorithm designed for this purpose is structured onto two main phases (see fig. 1).

A first Single Temporal Sample (STS) CD phase is oriented to detect both changes of a scene with respect to a background image and with respect to a close temporal frame in the sequence.

This phase aims at collecting information about long-term and short-term changes by means of a thresholding step on differences between the current image and the two reference images. In a latter data fusion phase, long-term and short-term information related to a limited temporal range immediately close to the considered time instant is organised into a feature vector by means of shift registers of limited length. The final decision whether a long-term change occurred into the scene is based on the content of such registers.

Let S={(i, j) : i = 1...M , j = 1...N } be a 2D lattice containing MxN pixels.
We define as $I_k = \{f_k(i,j) : (i,j) \in S\}$ the image belonging to the examined
sequence at the step k, k = 1...T, $f_k(i,j)$ being the value of the intensity response at
pixel (i, j) at time k. We refer to the background image as
BKG=$\{fb_k(i,j) : j \in S\}$.
Then, BG=$\{bg_k(i,j) = f_k(i,j) - fb_k(i,j) : (i,j) \in S\}$ represents the Simple
Difference (SD) image between the current image and the background image, while
FG=$\{fg_k(i,j) = f_k(i,j) - f_{k-1}(i,j) : j \in S\}$ is the result of the application of the
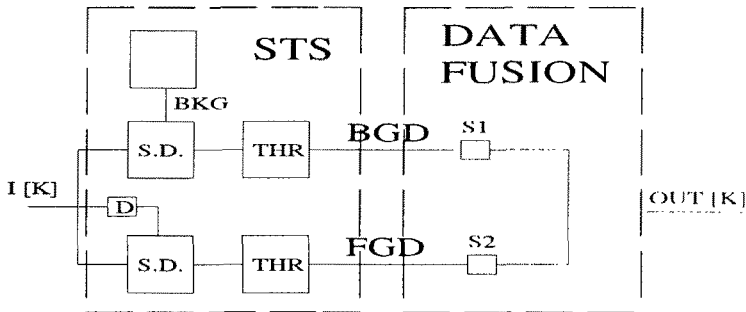simple difference algorithm to the current frame and the previous one.

Fig. 1. Block diagram of the Algorithm

Let us then define as

$$\bullet \quad BGD=\{ bgd_k(i,j) = \begin{cases} 1 & if \ \left|bg_k(i,j)\right| > th1 \\ 0 & if \ \left|bg_k(i,j)\right| < th1 \end{cases} : (i,j) \in S, bg_k \in BG \} \quad (1)$$

and as

$$\bullet \quad FGD=\{ fgd_k(i,j) = \begin{cases} 1 & if \ \left|fg_k(i,j)\right| > th2 \\ 0 & if \ \left|fg_k(i,j)\right| < th2 \end{cases} : (i,j) \in S, fg_k \in FG \} \quad (2)$$

the binary results derived from the application of the thresholding step to BG and
FG, respectively. The obtained data are collected into shift registers S1 and S2 of
length N and organised into a vectorial representation, which represents the feature
extraction step for the Data-Fusion CD module. S1 and S2 are characterised by the
following expressions:

$$\bullet \qquad S1 = \{fgd_i(i,j) : i = k - N + 1....,k , (i,j) \in S\} \qquad (3)$$

$$\bullet \qquad S2 = \{fgd_i(i,j) : i = k - N + 1....,k , (i,j) \in S\} \qquad (4)$$

Decision-making is performed on the basis of the content of S1 and S2 registers. In
particular, a thresholding step for the Data Fusion phase is performed on the
number $N_c(i,j)$ of couples (1,0) obtained by examining registers S1 and S2:

$$\bullet \quad N_c(i,j) = card \{(bgd_k(i,j) = 1, fgd_k(i,j) = 0) : i = k - N + 1....,k, (i,j) \in S\} \quad (5a)$$

In particular the output OUT(k) is obtained as:

$$\bullet \qquad \text{OUT}_k (i, j) = \begin{cases} 1 & \text{if } N_c(i,j) > ta \\ 0 & \text{otherwise} \end{cases} \qquad (5b)$$

Focusing the attention on couples (1,0) corresponds to discard from the decision of the presence of an abandoned object all those situations where a short-term change is detected. In fact, these situations can be more probably caused by occlusions (e.g., motion of people on the line of sight between the camera and the object itself) and, consequently, they do not carry useful information.

III. Probabilistic description

The problem on which this paper is focused is to determine, through a probabilistic criterion, the optimal value of the threshold *ta* considered in the Data Fusion Module to determine if a pixel belongs or not to a abandoned object.

The problem is stated as follows: we would like to find the output configuration which maximise the probability of correct detection P_D, or equivalently which minimise the error probability in the detection step. To this end, a cost function $J' = J'(P_{FA}, P_{MA})$ is defined, whose expression is the following:

$$\bullet \qquad J' = P_{MA} \, P_1 + P_{FA} \, P_0 \qquad (6)$$

where:

$$\bullet \qquad \begin{aligned} P_1 &= P(pixel \in \ abandoned \ object) \\ P_0 &= P(pixel \notin \ abandoned \ object) \end{aligned} \qquad (7)$$

and P_{FA} is given by:

$$\bullet \qquad P_{FA} = \int_{ta}^{+\infty} f(y/H_0)dy \qquad (8)$$

This is the probability that the algorithm decides for the presence of a abandoned object when we hypothesise (H_0) that the object is absent.
P_{MA} is expressed as:

$$\bullet \qquad P_{MA} = \int_{-\infty}^{ta} f(y/H_1)dy \qquad (9)$$

which is the probability that no object is detected when an abandoned object is present (H_1)
From the above expressions, the cost function J' is shown to depend on the threshold *ta*, i.e.:

$$\bullet \qquad J' = J'(ta, P_{FA}, P_{MA}) \qquad (10)$$

In the proposed model probabilistic occlusion model, the probability $P_{no}(i, j)$, called *probability of no-occlusions*, is introduced and it is shown that P_{FA} and P_{MA} depend on such a probability. $P_{no}(i, j)$ is defined as the probability that a pixel (i, j) is occluded by the presence on the line of sight of other objects in a given frame. Consequently, both P_{FA} and P_A will depend also on P_{no}, that is:

- $$J' = J'(\mathbf{P_{no}}, ta) \qquad (11)$$

The dependence of J' on $\mathbf{P_{no}}$ can be eliminated by considering for example an average case: in this case a cost function J can be computed by averaging J' on all possible values of $\mathbf{P_{no}}$, i.e.:

- $$J(ta) = \int J'(\mathbf{P_{no}}, ta)\, d\mathbf{P_{no}} \qquad (12)$$

This allows one to determine the optimal value of the threshold ta^* by minimising the cost function J that depends now only on ta, that is:

- $$ta^* : \arg \min_{ta} J(ta) \qquad (13)$$

IV. Optimization of *ta* threshold and Results

A. The algorithm

The problem illustrated in Section III has been solved by utilizing a bank of shift registers, one for each pixel of the image: in such a way the information about possible changes within the sequence is maintained in order to allow a "long-term" decision.

Let us analyse a set of images acquired at different temporal instants: f_{k+1} and f_k are the grey-level values for a pixel p, in the frame sequence, at the instants $k + 1$ and k, respectively; f_{b_k} (i,j) is the grey-level of the pixel p in the background image.

The information related to the difference frame-background is not sufficient in the case of an image sequence: we also need of an information about the temporal history of each pixel: for this purpose, we also need to check the variation of the luminosity function on the corresponding pixels in following images.

In order to make the system robust to the acquisition noise and to eventual occlusions, the decision if the pixel belongs or not to a left object is performed by exploiting the information about N frames previous to the considered one. This is done by constructing for each cell of the image two queues having N locations each and managed with a FIFO philosophy.(see fig. 2).
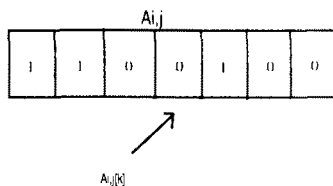

Fig. 2   Example of FIFO queue for the pixel (i,j).

The queues relevant to a single pixel p, i.e. S1 and S2, allow the memorisation of the registered information in N frames, being S1 the queue relative to the variations of the grey-levels between the frame the and background , having N locations, being S2 the queue relative to the variations of the grey-levels between following frames,

both having N locations, while S1[k] with k<N shows the content of the k-th location of the S1 queue.

In the frame-background queue S1 is then stored a value 1 if $\left| bgd_k(i,j) \right| > th1$, where $bgd_k(i,j) \in BGD$, while in the shift register S2 is stored the value 1 if $\left| fgd_k(i,j) \right| > th2$, where $fgd_k(i,j) \in FGD$.

By using the information contained in the shift-registers S1 and S2, it is now possible to calculate the figure $N_C(i,j)$ defined in (5b): if the number of (1,0) couples within the N frames is higher than the threshold ta, to the pixel of the output image is assigned the 1 value, corresponding to white grey-level in a binary image. The pixels belonging to the static areas are then these that in the image are white.

Let us recall that (1,0) are searched because it is possible to imagine that the presence of an abandoned object may cause a change in a frame with respect to the background image, while, since the object is static, there is no motion between two consecutive frames.

If there is a low threshold ta, it is sufficient a relatively low number of couples (1,0) in order to store a white color into the pixel (detection case). Therefore, for low values of ta, there will be high values of correct detection probability, but also of the false-alarm probability P_{FA}. Conversely, by increasing the threshold ta, the correct detection and the P_{FA} decrease. It is then necessary to find the threshold ta which maximise the correct detection probability and which minimise P_{FA}.

B. Quantitative results

Let us now to find the threshold ta in order to minimize the cost function J defined in (12): in all the experiments N=16 has been chosen, being N the length of the shift registers S1 and S2.

By using a training set image sequence $P_0 = 5*P_1$ has been put, i.e. the probability that a pixel don't belong to a left object has been fixed five times higher than the probability that belongs to the object itself, and the expression of the cost function J' which has been introduced in (6) has been computed and the graphical expressions of fig. 3 have been obtained: such graph depends on the P_{no} probability and each curve of the graph corresponds to a fixed value of the ta threshold.

By integrating the cost function in (6) over P_{no} with integration limits between 0 and 1 the graph of fig. 4 has been obtained. In this figure a value, that gives the integral of the cost function over P_{no} between 0 and 1, is represented for each integer value of ta.

In order to obtain the correct value of the optimal threshold ta^* we should have to consider the value which, given P_{no}, minimizes the cost function: the value of the threshold ta minimizes the integral of the cost function will be chosen, by then averaging the cost function itself over the no-occlusions probability P_{no}.

One can observe that the minimum value of the integral function is given by $ta \cdot 7$, which will represent the ta optimum value.
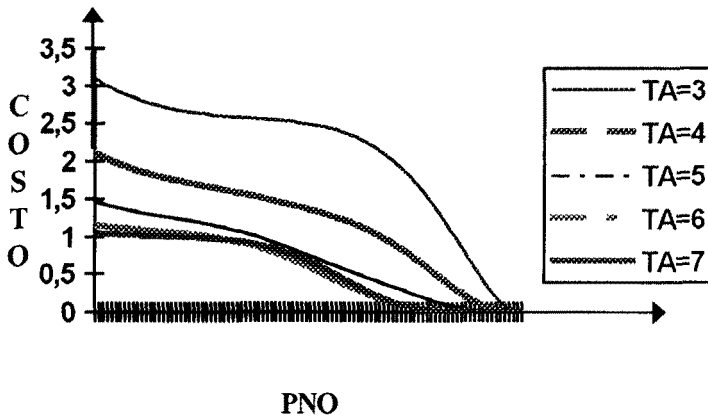

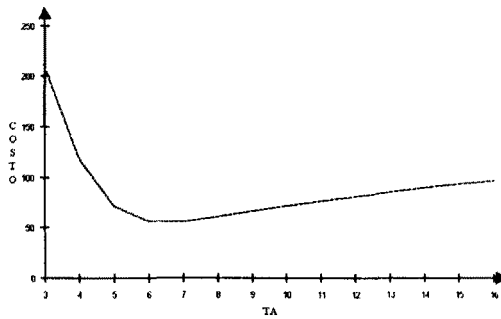Fig. 3 Cost function with parameter ta

Fig. 4 Integral of cost function with respect to ta threshold.

An example of the application of the proposed CD method to an image sequence of a surveilled waiting-room is now presented.

The test images were 256x256 grey-scale images (256 levels) and were extracted from an observed real sequence acquired in the surveilled waiting room of a railway station: fig. 5.a shows the background image of the waiting room; in fig. 5.b some abandoned objects are present: the proposed method aims to detect these objects themselves. Fig. 5.c is another frame of the same sequence in which some occlusions by other objects or persons are present and fig. 5.d shows the given result by applying the proposed method to a frame of the real sequence: in the last figure the abandoned objects are represented by the black pixel characterized by an high density.

We can see how the CD method is able to detect the static areas within the sequence (i.e. the abandoned objects) even in the presence of some occlusions: this is mainly due to the integration of two informative sources: the information coming from the frame-background difference and from the frame-by-frame one.

The method is also shown be able to evaluate and to discard other noise effects, like changes in illumination or noise in digitisation process.
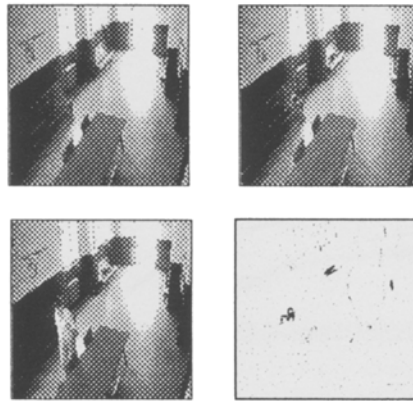

Fig. 6   a.b.c.d
Examples of the processed real sequence
and given result by the proposed method

The last remark is dedicated to the computation time that the proposed algorithm requires: in this case, that is by choosing N=16 and $ta=7$, the average computation time necessary to the CD algorithm to detect an abandoned object, measured on a Sun Sparc 20 Workstation, is equal to 2.3 sec.

Conclusions

A long-term CD method has been developed by means of the definition of a probabilistic model and the integration of two different informative sources. The model has been described and from a theoretical point of view and its real implementation by means of a shift-registers bank is presented. The algorithm is part of a surveillance system for unattended railway stations: results on a real image sequence confirm its validity.

References

[1] K. Skifstad, R. Jain, "Illumination Independent Change Detection for Real World Image Sequences", CVGIP, vol. 46, pp. 387-399, 1989.

[2] T.S. Huang (ed.), Image Sequence Analysis, Springer NY, 1981.

[3] G. Nicchiotti and E. Ottaviani, "Automatic vehicle counting form image sequences", in Proc. Of Int. Conf. On Time Varying Image Processing and Moving Object Recognition 1993, pp. 410-417.

[4] G.W. Donohoe, D.R. Husg and N. Ahmed, "Change detection for target detection and classification in video sequences", in Proc. IEEE Conf. Acoustics, Speech and Signal Processing, 1988, pp. 1084-1087.

[5] Z.S. Jain and Y.A. Chau, "Optimum Multisensor Data Fusion for Image Change Detection", IEEE Trans. On Systems, Man and Cybernetics, Vol. 25, No. 9, September 1995, pp. 1340-1347.

[6] P. Karmann, A.V. Brandt, "Moving Object Recognition using an adaptive background memory"" in Proc. on Int. Conf. On Time Varying and Moving Object Recognition 1990, pp. 289-296.