A Supervised Approach in Background Modelling for Visual Surveillance

P.Spagnolo, M.Leo, G.Attolico, A.Distante

Istituto di Studi sui Sistemi Intelligenti per l'Automazione - C.N.R.
Via Amendola 166/5, 70126 Bari, ITALY
{spagnolo,leo,attolico,distante}@ba.issia.cnr.it

Abstract: In this paper we address the context of visual surveillance in outdoor environments involving the detection of moving objects in the observed scene. In particular, a reliable foreground segmentation, based on a background subtraction approach, is explored. We firstly address the problem arising when small movements of background objects, as trees blowing in the wind, generate false alarms. We propose a background model that uses a supervised training for coping with these situations. In addition, in real outdoor scenes the continuous variations of lighting conditions determine unexpected intensity variations in the background model parameters. We propose a background updating algorithm that work on all the pixels in the background image, even if covered by a foreground object. The experiments have been performed on real image sequences acquired in a real archeological site.

1. Introduction

In the last years, motion detection has attracted great interest from computer vision researchers due to its promising applications in many areas, firstly visual surveillance. Due to a diffuse request of safety, surveillance systems are being installed in an increasing number of locations such as highways, streets, airports, homes, …. Our application context is the visual surveillance of archeological sites, where the main aim is to detect the presence of peoples and to recognize their gestures in order to timely identify illegal actions. In particular, the system should be able to recover the true shape of moving objects in order to allow a classifier to discriminate peoples from, for example, vehicles or animals. This paper concentrates only on the primary step of robust foreground segmentation of moving objects. The main aim is to introduce a technique that permits the correct extraction of moving objects and reduces the number of false alarms (as movements of vegetation caused by the wind). Furthermore, it is presented an updating algorithm that works on all the pixels in the background image, including the ones covered by foreground objects.

There are three conventional approaches to moving target detection: optical flow [1, 2, 3, 4]; temporal differencing [5]; and background subtraction [6, 7, 8]. Optical flow can be used to detect moving targets even in presence of camera motion, but the relative computational methods are very complex and cannot be coded into real-time algorithms without specialized hardware. Temporal differencing is very adaptive to

dynamic environments but, generally, does not allow all relevant features and the true shape of the moving objects to be obtained. Backgrounding methods, based on the thresholded difference of each image with respect to a model of the background scene, provide the most complete feature data and allow to recover the most reliable shapes of the moving objects. Their main drawback is the sensitiveness to dynamic scene changes due to lighting and extraneous events. In this paper we introduce a new backgrounding approach allowing a reliable moving foreground segmentation.

In literature, several approaches for automatically adapting a background model to dynamic scene variations have been proposed. Such methods differ mainly in the type of background model used and in the procedure used to update the model. Assuming that for most of the time the system processes a scene consisting of a relatively static situation, a normal distribution fitting slow changes in the scene by recursive updating has been proposed in [7] for modelling a completely stationary background. This approach models the background as a textured surface, each point of which is associated with a mean colour and a variance about that mean. This method uses a threshold for partitioning the background pixels into visible and occluded pixels. In each frame, the statistics of visible pixels are updated using a simple adaptive filter in order to compensate the background changes due to illumination and human motion. A background model handling also the small motions of background objects such as vegetation has been proposed in [6]; each pixel is represented by three values: its minimum and maximum intensity values, and the maximum intensity difference between consecutive frames observed during a training period. In these two methods, a simple threshold is used to determine whether a pixel belongs to background or foreground, however, this is not the case in real world where, normally, foreground and background cannot be set apart by a simple threshold. Therefore, many background subtraction algorithms consider noise explicitly into pixel measurement to extract moving objects. Noise measurements, rather than a simple threshold, are introduced in the statistical model proposed in [8, 9] where each pixel is represented using a running average and a standard deviation maintained by temporal filtering. In this way, the detection algorithm is more reliable than using a static threshold. Unfortunately, these approaches do not solve the problem of updating the regions covered by foreground objects.

Since the updating of each pixel in the background model pixel depends on the intensity variation with respect to its value in the current image alone, an error on labelling a foreground point as background point will determine an erroneous background model update. In our algorithm, this problems has been avoided using a stack-based updating algorithm: the results of motion detection in a certain number of frames are used to generate consistent values of the background parameters that will be used for the update. Another problem affecting the traditional backgrounding approaches is that they update the background model on the base of the variation evaluated at each single pixel. This implies that only the pixels corresponding to static points in the scene can be updated: that can be unacceptable if there are slowly moving objects in the scene. In fact, the intensity variation in correspondence with foreground points cannot be estimated, so in those regions the background model will remain unchanged as long as they are covered. This problem affects also the algorithms that cope with small movements in the background, as proposed in [6]. Instead, the proposed approach allows all the pixels in the image to be updated. The

basic idea is that the intensity variation of each pixel is estimated by integrating all the variations exhibited by the pixels with the same intensity value. In this way, also regions corresponding to moving background objects, as vegetation, can be updated.

In the rest of the paper, firstly a background model and the motion detection step are presented (section 2); then, an innovative approach for background updating, allowing the update of all the pixels, is proposed (section 3). Finally, the experimental results obtained on real image sequences acquired on an archeological site, and their comparison with the performance of other backgrounding methods, are reported (section 4).

2. Background subtraction

Foreground object segmentation is a primary and fundamental step of visual surveillance systems: the results of this step are the inputs for the subsequent processing (object recognition, motion analysis, …). So it is very important to extract correctly the moving objects. In outdoor environments this is not so easy, because of the variations in lighting conditions and of the small movements of background objects, as trees blowing in the wind. That makes necessary to develop very reliable motion detection algorithms, that should be adaptive to luminance variations and able to reduce the number of false alarms due to movements without interest for the surveillance task. A supervised algorithm for background modeling has been implemented for dealing with these constraints. The proposed approach uses a training sequence of frames in which there are only 'legal' moving objects, i.e. trees, and not human figures. The system uses this training period to determine a series of parameters which contain the information about tolerable motions. So during the normal surveillance operations, when a movement is detected it is considered legal if it is similar to the movements registered during the training period, while it is classified as illegal otherwise and further surveillance steps, such as object recognition, are activated. To implement this idea, each pixel has been modeled through statistical information: the evolution of the intensity value at each pixel during the training period is used to calculate mean and variance at that pixel; the variance is directly correlated with the nature of the corresponding point: a generic background point will probably have a small variance, while a point close to a moving object (such as a tree) will have higher variance value, due to a greater dispersion of the intensity values registered during the training period.

This kind of approach is not sufficient for dealing with the presence of small movements of background objects. The simple information about mean and variance must be integrated with further data about legal motions: illegal movements, i.e. a walking person, could generate variations from the mean lower than the variance, being erroneously classified as parts of the background. It is necessary to introduce a further parameter, containing the information about admissible values for each point at time t as a function of the values registered at time $t-1$. An approach based on a a-posteriori probability has been implemented for estimating admissible values. During the supervised training, the a-posteriori probability is calculated at each pixel:

$$P(I_t(x,y) = j \mid I_{t-1}(x,y) = i) \quad (1)$$

as the probability that the pixel (x,y) could have an intensity value j at time t if its value at time $t-1$ was i. This a-posteriori probability is calculated and memorized as an hash table for each pixel, correlating intensity values at time *t* with the ones at time *t-1*. During the working phase, if the difference between intensity values at subsequent times is greater then the correspondent value in the hash table, the corresponding pixel is marked as 'foreground'. Because implementing an hash table of (256 x 2) elements for each pixel is too expensive in terms of both computational time for the updating and in terms of memory resource occupancy (for an image of 528 x 512, it could be used 528*512*256*2=132 Mb!) a very good trade-off has been obtained storing, for each pixel, only the maximum of all admissible values (528*512=264 Kb). In this way, a pixel is marked as 'foreground' not only if the difference between its intensity values at time *t* differs from its mean by more than the variance, but also if the difference between its current and previous values is greater than the maximum variation admissible for that pixel.

In other words, the proposed motion detection algorithm marks a pixel as a foreground pixel if, at time *t*:

$$\left|I_t(x,y) - m(x,y)\right| > s\,(x,y) \quad (2)$$

or

$$\left|I_t(x,y) - I_{t-1}(x,y)\right| > P(x,y) \quad (3)$$

where $I_t(x,y)$ and $I_{t-1}(x,y)$ are the intensity values for the pixel (x,y) at time *t* and *t-1*, $m(x,y)$ and $s\,(x,y)$ are mean and variance respectively of the intensity values observed during the supervised training period, and $P(x,y)$ is the maximum difference between intensity values that are consecutive in time, observed during the whole training period:

$$P(x,y) = \max_{t \in T} \left|I_t(x,y) - I_{t-1}(x,y)\right|, \quad \text{T=frames of the training set} \quad (4)$$

After this step, in the resulting binary image there are many small clusters of pixels that must be removed: a one step filter removes blob whose size is lower than a certain threshold. Finally we obtain an image with only foreground objects, each of whose has been extracted with its own shadow. Because the presence of shadows changes the shape of objects in an unpredictable way, causing serious trouble to the following step object recognition, the algorithm proposed in [10] has been used for removing shadows. It starts from the assumption that a shadow is an abnormal illumination of a part of an image due to the interposition of an opaque object with respect to a bright point-like illumination source. From this assumption, we can note that shadows move with their own objects but also that they have not a fixed texture, as real objects: they are half-transparent regions which retain the representation of the underlying background surface pattern. Therefore, our aim is to examine the parts of the image that have been detected as moving regions from the previous segmentation

step but with a structure substantially unchanged with respect to the corresponding background. To do this, firstly a segmentation procedure has been applied to recover large regions characterized by a constant photometric gain; then, for each segment previously detected, the correlation between pixels is calculated, and it is compared with the same value calculated in the background image: segments whose correlation is not substantially changed are marked as shadow regions and removed. So, the final image contains only motion objects without shadows: these are the input for the object recognizer described in [11].

3. Background updating

Any background subtraction approach is sensitive to variations of the illumination; to solve this problem the background model must be updated. Traditional updating algorithms have a serious problem: they operate the updating only on the pixels that have been labeled as 'background' in the last frames. If in a region there is a moving object, the corresponding background pixels are left unchanged. In particular in presence of slow moving objects, as a person staying in a certain region for a certain period of time, this can invalidate the results. In addition, erroneously labeling foreground and background points could determine a wrong update of the background model. The proposed approach allows all the pixels of the background to be updated, even if they correspond to points that at time *t* are masked by foreground objects: every background pixel can be updated even if currently invisible. In literature, many approaches update the background at each frame; in our case, the implemented surveillance system works at a frame rate of 30 Hz, so it is not necessary to update the background at each frame because relevant variations in a very short time are not probable. The proposed approach, according with the background model implemented, consists in a stack-updating: the number of frames to be used for calculating the new values of mean and variance is fixed (i.e. 80, 100 or 150 frames). During this period, new values of the statistical parameters are calculated in the static points of the image. The updating rule (4) allows a parameter α, tacking values in the range (0,...,1), to control the updating process by weighting differently the two terms:

$$m_{n+1}(x,y) = \alpha * m_n(x,y) + (1-\alpha) * m_{n-1}(x,y) \quad (5)$$

$$s_{n+1}(x,y) = \alpha * s_n(x,y) + (1-\alpha) * s_{n-1}(x,y) \quad (6)$$

$$P_{n+1}(x,y) = \alpha * P_n(x,y) + (1-\alpha) * P_{n-1}(x,y) \quad (7)$$

where the *n+1* indicates the new updated values, *n* indicates the values calculated during the last observation period, and *n-1* indicates the old values of the parameters. The updating procedure described above must be done only for pixels that have been classified as static for most of the observation period (i.e. 80%). Because the continuous variations of the lighting conditions in outdoor environment determine uniform variations in all the image intensity values, it is possible to update all pixels in the image. The idea is that pixels with the same mean of intensity values will assume the same value after updating. So, it is possible to update pixels covered by

foreground objects by simply calculating the updating average of the pixels with the same intensity value. Therefore the updating value relative to each pixel of the background model covered by a foreground object is estimated by averaging all the different values $m_n(x,y)$ exhibited by all the pixels $\{(x,y)\}$ with the same intensity value $m_{n-1}(x,y) = b_i$.

$$m(b_i) = \frac{1}{N(b_i)} \sum_{\{(x,y) \in I^t | m_{n-1}(x,y) = b_i\}} m_n(x,y) \qquad (8)$$

where $\{b_i\}_{i=1,...,n}$ are the n different intensity values that each pixel can assume, and $N(b_i)$ is the number of pixels in the background model with intensity mean value b_i.

So, the update rule for the mean will be:

$$m_{n+1}(x,y) = a * m(m_{n-1}(x,y)) + (1-a) * m_{n-1}(x,y) \qquad (9)$$

In an analogous way σ and P can be updated. Moreover, in order to avoid the accumulation of the error over time, the background model is periodically reinitialized in all the regions labeled as static for a long time interval. In this case, α can be set to very low values allowing the system to self-reinitialize.

4. Experimental results

The experiments have been performed on real image sequences acquired with a static TV camera Dalsa CA-D6 with 528 X 512 pixels; the frame rate selected is 30Hz. The processing is performed with a Pentium IV, with 1,5 GHz and 128 Mb of RAM. The characteristics of each test sequence are resumed in the table 1. The sequences are acquired in a real archeological site while people were simulating the movements normally performed by intruders.

Table 1 The test sequences

Sequence number	Frames	Number of frames of the training	Number of frames of the updating-stack
1	894	200	150
2	387	150	100
3	1058	200	150
4	551	150	100

The results obtained applying the proposed motion detection algorithm are very encouraging. In the following images the obtained results are compared with the one provided by two very common motion detection approaches. In particular, the first column of fig. 1 shows two images of a sequence; the second column illustrates the results obtained applying a statistical background model like that proposed in [8]: it is evident that the presence of moving small trees heavily affects the correct extraction of shapes, a problem clearly mentioned in the relative paper. The third column shows the results obtained implementing the approach described in [6]: it can be seen that

the moving trees are not detected by the system, but the region where the background model is obsolete due to the presence of a human that moves slowly, the result is affected by a large noise. Finally, the last column depicts the results obtained using the proposed approach: moving trees are not detected and the quality of the resulting shapes is higher then the previous ones, even in regions where people stayed for a quite long period of time.

Fig. 1: The comparison between the results obtained applying to a pair of images (1st column) a traditional statistic background modelling (2nd column), a supervised approach using a traditional background update (3rd column) and the proposed algorithm (4th column).

5. Conclusions and Future Works

This work deals with the problem of outdoor motion detection in the context of video surveillance. A supervised approach for background subtraction has been implemented to reduce the amount of false alarms caused by small movements of background objects. A new updating algorithm has been implemented to allows the update of all the background pixels, even if covered by foreground objects. The experiments show that the proposed approach works better then other similar techniques proposed in literature, in particular in presence of slowly moving objects. Future work will investigate the improvement of the training algorithm, using a non supervised approach. This is a very important requisite for the real implementation on a visual surveillance system, that needs to be very reliable. Another important improvement of the system is expected by a more correct detection of background object movements even if they exhibit intensity changes greater than the ones observed during the training period. All these goals must be obtained by reducing the dependency of the algorithm on the conditions occurred during the training period.

References

[1] S. Fejes, L.S. Davis, Detection of independent motion using directional motion estimation, Technical Report. CAR-TR-866, CS-TR 3815, Univ. of Mar., Aug. 1997.

[2] S. Fejes, L.S. Davis, What can projections of flow fields tell us about the visual motion, In Proc Intern. Confer. on Computer Vision ICCV98, 1998, pp. 979-986.

[3] S. Fejes, L.S. Davis, Exploring visual motion using projections of flow fields, In. Proc. of the DARPA Image Underst. Work., pp.113-122, New Orleans, LA,1997.

[4] L.Wixson, M.Hansen, Detecting salient motion by accumulating directional-consistent flow, In proc. of Intern. Conf. on Comp. Vis., 1999, vol II, pp 797-804.

[5] C.Anderson, P.Burt, G. Van Der Wal, Change detection and tracking using pyramid transformation techniques, In Proc. of SPIE – Intell. Robots and Comp. Vision Vol. 579, pp.72-78, 1985.

[6] I.Haritaoglu, D.Harwood, L.Davis, A Fast Background Scene Modeling and Maintenance for Outdoor Surveillance, ICPR, pp.179-183, Barcelona,2000.

[7] C.Wren, A.Azarbayejani, T.Darrell, A.Pentland, Pfinder: Real-time tracking of the human body, IEEE Trans. on Patt. An. and Mach. Intell. 19(7): pp.780-785, 1997.

[8] T.Kanade, T.Collins, A.Lipton, Advances in Cooperative Multi-Sensor Video Surveillance, Darpa Image Underst. Work., Morgan Kaufmann, Nov. 1998, pp. 3-24.

[9] H. Fujiyoshi, A. J. Lipton, Real-time human motion analysis by image skeletonisation, IEEE WACV, Princeton NJ, October 1998, pp.15-21.

[10] P. Spagnolo, A. Branca, G. Attolico, A. Distante: Fast Background Modeling and Shadow Removing for Outdoor Surveillance, IASTED VIIP,2002, pag. 668-671.

[11] M. Leo, G. Attolico, A. Branca, A. Distante: Object classification with multiresolution wavelet decomposition , in Proc. of SPIE Aerosense 2002, conference on Wavelet Applications, 1-5 April, 2002, Orlando, Florida, USA.