ROBUST MOTION DETECTOR FOR VIDEO SURVEILLANCE APPLICATIONS

Antonio Albiol, Cristina Sandoval, Alberto Albiol, Valery Naranjo and José Manuel Mossi

Departamento de Comunicaciones, Universidad Politécnica de Valencia
Camino de Vera s/n
46022 Valencia, Spain

ABSTRACT

[1] This paper presents a robust motion-detector video sensor. It is intended to operate in surveillance applications for long periods of time with time-varying noise level. It makes use of the fact that whenever there is no motion a similarity measure between frames tends to have similar values.

1. INTRODUCTION

One of the facts that is changing our world is the ubiquitous use of cameras for surveillance. These cameras generate huge amounts of raw information that is either stored or processed by humans. Processing of this visual information by humans is normally a tedious work since most of the time the raw information has no interest.

These days, video-based surveillance systems are migrating from the conventional analog CCTV solutions to the so called third generation surveillance systems [1], which stores and transmits digital video. An interesting application area of video processing is the development of *video sensors*. A video sensor is nothing but a digital video analysis tool that extracts meaningful information from the raw video. Several advantages arise from the use of video sensors, namely constant performance, cost and privacy protection.

Some difficulties that can be found when developing successful video sensors are:

- Real world images; this refers to the fact that video sensors must operate on almost unrestricted conditions of lighting, environments, and cameras quality.
- Easy setup and operation; this is crucial in order to be able to produce something that is accepted by the surveillance industry. A CCTV installer should be able to set up a good video sensor, without critical parameter tuning or tedious trainings.
- Operation during long periods in changing environmental conditions.
- Low computational cost. A single computer must deal usually with various video streams and also the computational power must be shared with other tasks such as video compression.

One key advantage is that a video-sensor may work in cooperation with a human operator. In this case the automatic video sensor behaves like a prefilter of candidate interesting situations that need further evaluation by the human. This way, the human is faced only to the (more likely) relevant information.

This paper deals with an apparently simple but very useful video sensor. Its purpose is to detect when a change occurs in the scene. Its application is fully justified since normally static video is of no interest; it can be used to trigger recording or as an alarm generator.

Our sensor falls in the category of change detectors. Different approaches have been reported in the literature depending on the specific needs. In [2] different change detectors are presented, and some of them have the property of being partially able to ignore changes due to illumination. In [3] a scheme that also ignores illumination changes is presented; in this case the input are JPEG compressed images and there is no need of full decompression. The technique proposed in this paper makes no attempt to ignore changes due to illumination changes; in other words we will find also the changes due to true motion and the changes due to sudden illumination changes. Other papers deal with the segmentation of moving portions of the image ([4], [5] [6], [7]). However we are interested in finding if there is any change in a certain region. In other words, we are interested in a global indication of the presence of change. This is the same case as in [8]. The kind of robustness that we are interested in is against changes in the noise level. Variations

This work has been funded by Spanish Minsitry of Science and Technology under grant TIC 2002-02469.

in noise level occur mainly due to the presence of AGC in most CCTV cameras.

The rest of the paper is organized as follows. Section 2 presents the proposed technique. Section 3 presents a short study on the choice of the parameters of our algorithm. Finally, results obtained with the use of the proposed video sensor are presented.

2. ALGORITHM DESCRIPTION

The motion detection can be applied on the whole image or on a region of interest (ROI) that we will subsequently denote as $\mathcal{R}$. The number of pixels of $\mathcal{R}$ will be denoted as N. Let's denote as $I(t,p)$ the input image at time instant t and position p. Then we obtain a simple measure of similarity between $\mathcal{R}$ at t and $t-1$ as follows [9]:

$$D(t) = \frac{1}{N}\sum_{p\in\mathcal{R}} |I(t,p) - I(t-1,p)| \tag{1}$$

In absence of motion, we would ideally have

$$I(t,p) = I(t-1,p) \tag{2}$$

and then $D(t) = 0$. However noise is always present in images, and a better model of the images in the absence of motion would be

$$I(t,p) = I(t-1,p) + n(p) \tag{3}$$

where $n(p)$ is a zero mean noise, with variance σ_n^2. In this, case $D(t)$ is a random process with the following statistics:

$$\bar{D} = k_1\sigma_n \tag{4}$$

$$\sigma_D^2 = k_2\frac{\sigma_n^2}{N} \tag{5}$$

where constants k_1 and k_2 are constants that depend on the shape of the probability density function of $n(p)$. For $n(p)$ Gaussian,

$$k_1 = \sqrt{\frac{2}{\pi}} = 0.7979 \qquad k_2 = \frac{\pi-2}{\pi} = 0.3634$$

Thresholding the value of $D(t)$ could be used to detect motion. The value of the threshold should be set according to the noise level σ_n^2:

$$T = \bar{D} + k_3\sqrt{\sigma_D^2} = \sigma_n\left(k_1 + k_3\sqrt{\frac{k_2}{N}}\right) \tag{6}$$

where k_3 depends on the desired probability of false alarm. Notice that, regardless of the shape of the probability density function of $n(t)$, in absence of motion, $D(t)$ is almost Gaussian for N large enough due to the central limit theorem.

In practice the value of σ_n^2 is unknown and should be estimated *when no motion* is present. This could be done at the setup stage. However, this approach would have two main drawbacks:

1. We must be sure that there is no motion when the sensor is started.
2. The noise level depends on the illumination. It normally increases when illumination decreases.

This facts cause that a continuous re-estimation of the threshold is necessary.

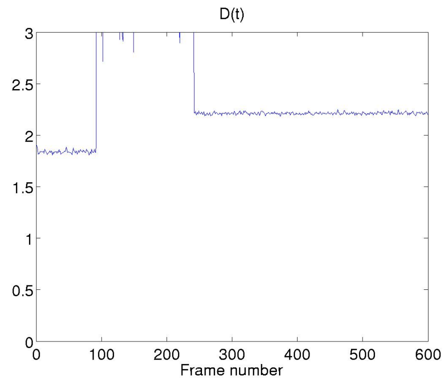

Fig. 1. Person leaving a room and switching off the lights. Values higher than 3 have been clipped in the plot.

In Figure 1 we can see an example of $D(t)$. In this example the camera was pointing to a door. Then someone passes, switches off the room light and leaves. Notice that the mean level of $D(t)$ has changed after the illumination change.

One possible strategy to update the threshold would be to make a recursive update *when no motion is present*. Notice that we need to know that there is no motion to update the threshold T that indicates if there is motion or not.

In the example of Figure 1, if we had fixed the threshold around a value of 2 (very low probability of false alarm), we would continue deciding that there is motion after the *true motion*, since σ_n^2 has increased after switching off the lights, and then would not update the threshold until the original lighting conditions were reestablished (perhaps hours later!!). It could be argued that setting a higher threshold T would have solved the problem; the question is how much higher?. Raising it in excess would lower the sensitivity of the video sensor, producing missed detections of motion. Figure 1 shows that in the motion interval, we find values smaller than 3.

The main idea proposed by this paper is to use the fact that when no motion is present, $D(t)$ has a small variance. This fact can be used to determine when to update threshold T. The reason why variance of $D(t)$ is small when there is no motion is that in equation 5, N, the number of pixels of the region, is normally large.

The variance can estimated by time averaging $D(t)$ and $D^2(t)$, using a linear low pass filter $H(z)$ with unity gain for DC. We will subsequently denote as $Var_D(t)$ the local variance of $D(t)$. Figure 2 shows the block diagram that we have used to estimate $Var_D(t)$. The filter $H(z)$ that we have used is:

$$H(z) = \frac{1-a}{1-az^{-1}} \quad (7)$$

which has one single parameter, a. A discussion on the choice of a will be presented in section 3.

Figure 3 shows a semi-log plot of $Var_D(t)$, for the same interval as in Figure 1. It can be seen that direct thresholding of $Var_D(t)$ would allow us to determine the presence of motion.

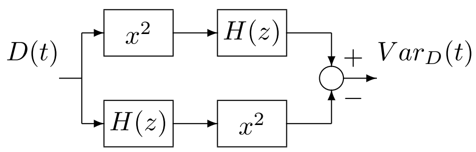

Fig. 2. Block diagram used to estimate $Var_D(t)$.

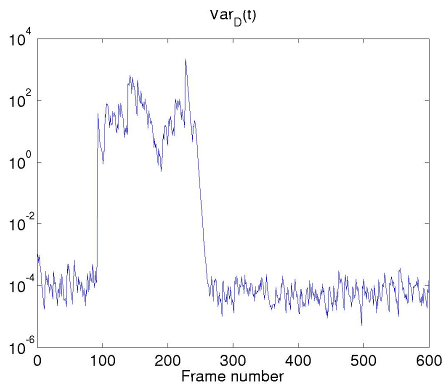

Fig. 3. Person leaving a room and switching off the lights. Evolution of local variance.

3. PARAMETERS ADJUSTMENT

In order to detect motion using variance we have two parameters to adjust:

- The filter parameter a. It controls the duration of the impulse response of the filter and determines the effective number of samples involved in the variance computation.
- The threshold on the variance T_V

3.1. Filter parameter adjustment

For the filter parameter, we wish two properties:

1. It must provide a fast response.
2. It must have a large sensitivity.

To determine the optimum value of a for a fast response, we will consider the following sequence of $D(t)$ values:

$$D(t) = \begin{cases} C & t < t_0 \\ C+1 & t = t_0 \end{cases} \quad (8)$$

This models a sudden increase in the value of $D(t)$. We would like to find the value of the filter parameter a that maximizes $Var_D(t_0)$. It is easy to show that if $D(t)$ is as in equation 8, then

$$Var_D(t_0) = a - a^2 \quad (9)$$

The maximum is reached for $a = 0.5$.

The use of variance of $D(t)$ for detecting motion is based on the fact that when there is motion the value of $D(t)$ exhibits larger variations that when there is no motion. For the second desired property, let us consider:

$$D(t) = \begin{cases} C & t \text{ odd} \\ C+1 & t \text{ even} \end{cases} \quad (10)$$

as a test sequence for $D(t)$, and will try to find which value of a maximizes the $Var_D(t)$. It is easy to show that in this case

$$Var_D(t) = 0.25\left(1 - \left(\frac{1-a}{1+a}\right)^2\right) \quad (11)$$

which is maximum for $a = 1$.

The above results tell us that a compromise value must be chosen between 0.5 and 1 to fulfill the two requirements. We can determine the variation with a of the two merit criteria with respect to their corresponding maximum value.

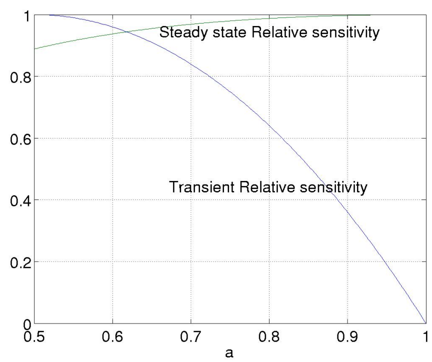

Fig. 4. Relative variation of the two criteria for choosing a.

This variation is shown in Figure 4. The crossing point is at:

$$a_{opt} = \frac{\sqrt{5}-1}{2} = 0.618$$

With this choice the value of any of the two criteria is 94.4% of the maximum.

3.2. Variance threshold

Thresholding $Var_D(t)$ can be used to determine the existence of motion. The rationale behind this assertion, is that quantity $D(t)$ may be larger or smaller depending on the noise level, but as long as the number of pixels, N, in the region $\mathcal{R}$ is large, $D(t)$ will be quite constant, or in other words will have a small variance. On the other hand, $D(t)$ tends to have quite different values when motion is present.

Assume that we have an upper bound for σ_n^2. Let's call this upper bound $\sigma_{\max}^2$. Then, the mean value in absence of motion for $Var_D(t)$ is (see equation 5):

$$\overline{Var_D(t)} < \sigma_{\max}^2 \frac{k_2}{N} \tag{12}$$

The variance of $Var_D(t)$ depends on fourth order moments of $n(t)$. Rather than trying to use these moments, we have found that a threshold about ten times larger than the value in equation 12 yields virtually no false alarm and allows a high detection rate.

We have also observed that for a large range of video cameras, and considering illuminations from complete darkness to standard lighting conditions for the camera, a general value for $\sigma_{\max}^2$ can be obtained. Using this value will give us a practical threshold:

$$T_V = \frac{500}{N} \tag{13}$$

Choosing this threshold for the variance, *eliminates the need of adjusting any parameter* for proper video sensor operation.

4. PERFORMANCE EVALUATION

In this section we will present the results that we have obtained with the motion video detector presented in the previous sections. We have used the motion detector to trigger recording. We have evaluated the false alarm rate by looking at the recorded video and checking if there was true motion at each event. Preliminary results have shown no false alarm in one week , at three different locations. The images were acquired continuously (two images per second,24 hours a day) in lighting conditions ranging form complete darkness, to natural lighting including artificial lights. The probability of detection results have also been good. The detection ability depends on the ratio between the number of pixels actually changing in $\mathcal{R}$ and the total number of pixels in $\mathcal{R}$, N.

5. CONCLUSIONS

This paper has presented a motion detector based on the analysis of images from a video sequence. The main idea is, rather than measuring the similarity between consecutive frame pairs, we exploit the fact that when there is not motion, similarity measures tend to be very similar along time, while when there is motion these measures exhibit large variations.

The proposed technique is very simple to install and operate by non-specialized operators, and needs no parameter adjustment.

6. REFERENCES

[1] C.S. Regazzoni, V. Ramesh, and G.L. Foresti, "Special issue on video communications, processing and understanding for third generation surveillance systems.," *Proceedings IEEE*, vol. 89, no. 10, 2001.

[2] E. Duracan and T.Ebrahimi, "Change detection and background extraction by linear algebra," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1382–1402, October 2001.

[3] Julio Pons, Josep Prades-Nebot, Antonio Albiol, and Jesus Molina, "Fast motion detection in the compressed domain for video surveillance," *IEE Electronics Letters*, vol. 38, no. 9, pp. 409–411, April 2002.

[4] Hieu T. Nguyen, Marcek Worring, and Anuj Dev, "Detection of moving objects in video using a robust motion similarity measure," *IEEE Transaction on Image Processing*, vol. 9, no. 1, pp. 137–141, 2000.

[5] Christophe Dumontier, Franck Luthon, and Jean-Pierre Charras, "Real-time dsp implementation for mrf-based video motion detection," *IEEE Transactions on Image Processing*, vol. 8, no. 10, pp. 1341–1347, October 1999.

[6] Kristine Matthews and Nader Mamazi, "A bayesian approach to uncovered background and moving pixel detection," in *ICASSP 1995*. IEEE, 1995, pp. 2245–2248.

[7] Julian F.Y. Cheung, Michael C. Wicks, Gerard J. Genello, and Ludwik Kurz, "A statistical theory for optimal detection of moving objects in variable corruptive noise," *IEEE Transactions on Image Processing*, vol. 8, no. 12, pp. 1772–1787, 1999.

[8] Stewart Young, Michael Forshaw, and Mark Hodgetts, "Change detection for perimeter surveillance," Internal report, Sira Technology Centre,, 1999.

[9] Ramesh Jain, Rangachar Kasturi, and Brian G. Schunk, *Machine Vision*, Mc Graw Hill, 1995, ISBN 0-07-032018-7.