

# Small Moving Object Detection in Video Sequences

Rabi Zaibi<sup>1</sup>, A. Enis Çetin<sup>1</sup>, Yasemin Yardımcı<sup>2</sup>

<sup>1</sup> Department of Electrical and Electronics Engineering,  
Bilkent University, Ankara, Turkey

<sup>2</sup> Middle East Technical University, Ankara, Turkey  
E-mail: cetin@ee.bilkent.edu.tr

## Abstract

In this paper, we propose a method for detection of small moving objects in video. We first eliminate the camera motion using motion compensation. We then use an adaptive predictor to estimate the current pixel using neighboring pixels in the motion compensated image and, in this way, obtain a residual error image. Small moving objects appear as outliers in the residual image and are detected using a statistical Gaussianity detection test based on higher order statistics. It turns out that in general, the distribution of the residual error image pixels is almost Gaussian. On the other hand, the distribution of the pixels in the residual image deviates from Gaussianity in the existence of outliers. Simulation examples are presented.

## 1 Introduction

Detection of small moving objects can be a complicated task when there is noise and the video camera is in motion. Classical object detection methods [1], [2] are geared for large objects with clear features and boundaries whereas in our problem the moving region or object may consist of only a few pixels.

In our method, we first eliminate the camera motion using motion compensation. After motion compensation, the resulting image basically contains the moving regions and objects. We process this image using two-dimensional (2-D) adaptive filter. We adaptively predict each pixel using neighboring pixels

and obtain a residual error image. It turns out that the distribution of the residual error image pixels is almost Gaussian in general. However, small moving objects produce outliers in the residual image as the pixels of the small moving objects cannot be predicted using the neighboring pixels. We detect the outliers using Higher Order Statistical (HOS) based Gaussianity test [4], [5]. In static regions the test statistics is very close to zero whereas in regions containing the moving object, the distribution of pixels deviate from Gaussianity and test statistic produces large values.

In Section 2, we present the 2-D adaptive filtering method which removes the static background. In Section 3, we review the HOS based statistical tests for moving object detection and present the results of simulation studies in Section 4.

## 2 2-D Adaptive Filtering

In Figure 1, the adaptive filtering scheme in two dimensions is illustrated. In the adaptive filtering process, an image pixel  $x[m, n]$  at location  $(m, n)$  is predicted as a weighted average of pixels in its region of support. The region of support,  $\mathcal{R}$ , of the adaptive filter is chosen as the pixels surrounding the pixel to be predicted as shown in Figure 1. The predicted pixel value  $\hat{x}[m, n]$  is given as  $\hat{x}[m, n] =$

$$\sum_{\substack{k=-n_1 \\ (k,l) \neq (0,0)}}^{n_1} \sum_{l=-n_2}^{n_2} w_{(m,n)}[k,l]x[m-k, n-l],$$

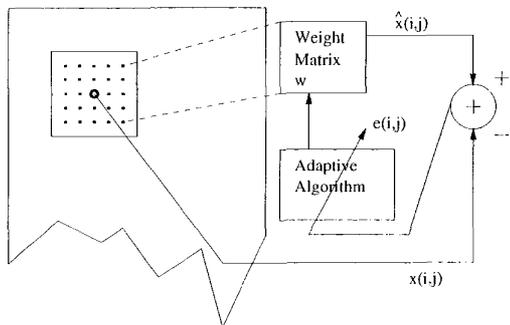


Figure 1: 2-D adaptive filter structure

$$m = 0, \dots, N_1 - 1, \quad n = 0, \dots, N_2 - 1 \quad (1)$$

where  $x$  is the motion compensated video frame of size  $N_1 \times N_2$ ,  $w_{(m,n)}$  are the weight values at  $(m, n)$ , and  $(2n_1 + 1) \times (2n_2 + 1)$  is the size of the region of support,  $\mathcal{R}$ , of the adaptive filter. In our case,  $n_1 = n_2 = 1$ .

The prediction error of the adaptive filter at location  $(m, n)$  is calculated as

$$e[m, n] = x[m, n] - \hat{x}[m, n] \quad (2)$$

The weight values  $w_{(m,n)}[k, l]$  are adapted according to the 2-D LMS-type adaptation algorithm [8]:

$$w_{(m+1,n)}[k, l] = w_{(m,n)}[k, l] + \mu \times e[m, n] \times x[k, l] \quad (3)$$

where  $(k, l) \in \mathcal{R}$ , and  $\mu$  is the adaptation constant. These weights are adapted using Equation 3 while processing the image in the horizontal direction. In a similar way, in the vertical direction the weight  $w_{(m,n+1)}[k, l]$  replaces  $w_{(m+1,n)}[k, l]$  in Equation 3. In Figure 2 (a) and (b) a frame containing a moving object and its following frame are shown. The image obtained from motion compensation is filtered using the 2-D adaptive prediction filter and the resulting residual error image is shown in Figure 2 (b). The moving object detection is then performed on the error image  $e[m, n]$ .

### 3 Higher Order Statistical Test

It has been experimentally observed that in regions with no moving objects, the error image  $e[m, n]$

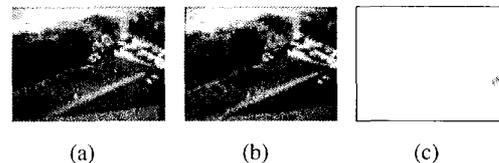


Figure 2: (a) A frame containing a moving object, (b) the following frame, (c) the error image after 2-D adaptive filtering (inverted).

has Gaussian like distribution, whereas in regions containing small moving objects it contains outliers. This is due to the fact that pixels of a small moving object cannot be accurately predicted using the surrounding pixels.

Higher order statistical tests are successfully used in the detection of microcalcifications in mammogram images [5], and detecting objects in noisy images [6]. In this paper, we use a Gaussianity test developed in [5] and [4]. The higher order statistic  $h(I_1, I_2, I_3, I_4)$  is based on the sample estimates of the first four moments  $I_1, I_2, I_3, I_4$  of the prediction error [4]. Estimates of the moments are given by

$$I_k = \frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N e^k[m, n] \quad (4)$$

where  $e[m, n]$  represents the error value at location  $(m, n)$  calculated in Equation 2 and  $M \times N$  is the size of the region in which  $I_k$  is estimated. The statistic  $h(I_1, I_2, I_3, I_4)$  is defined as follows:

$$h(I_1, I_2, I_3, I_4) = I_3 + I_4 - 3I_1(I_2 - I_1^2) - 3I_2^2 - I_1^3 + 2I_1^4 \quad (5)$$

It is ideally equal to zero when the distribution is Gaussian. It takes large values when the underlying distribution deviates from Gaussianity. Outliers in the error image are mainly due to moving objects and  $h(I_1, I_2, I_3, I_4)$  takes large values in such regions containing outliers.

In our approach, the image is analyzed block by block. The statistic is calculated within each block inside the image. These blocks may overlap as shown in Figure 3. In our experimental work we used blocks of size  $M = 15$  by  $N = 15$  where overlapping occurs at 3 pixel steps.

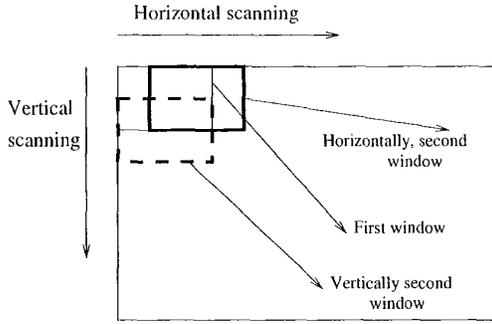


Figure 3: Illustration of overlapping windows.

The detection method can be considered as a hypothesis testing problem in which the null hypothesis  $H_0$  corresponds to the no moving object case and  $H_1$  corresponds to the presence of a moving object:

- $H_0: |h(I_1, I_2, I_3, I_4)| < T_h$
- $H_1: |h(I_1, I_2, I_3, I_4)| \geq T_h$

The threshold  $T_h$  is experimentally determined. The blocks in which the test statistic exceeds the threshold,  $T_h$ , are marked as regions containing the small moving objects.

## 4 Experimental Results and Conclusion

In this section, we present simulation studies. We test the performance of the detection scheme by analyzing 27 video sequences containing small moving objects on various backgrounds. As described in Section 1, motion compensated images are obtained in the first step. A classical block matching based motion compensation algorithm with subpixel accuracy is used [2].

In the second step, motion compensated images are filtered using the adaptive predictor and the residual error images are obtained. The values of the test statistic  $h$  in 12 video sequences are given in Table 1. It is clear from this table that a threshold can be selected which can distinguish moving objects from the background.

| Regions               | Minimum | Maximum |
|-----------------------|---------|---------|
| With moving object    | 2.3     | 7.5     |
| Without moving object | -0.41   | 0.5     |

Table 1: Values of the test statistic  $h(I_1, I_2, I_3, I_4)$  in regions with and without moving objects.

| Algorithms          | False Alarms | Miss |
|---------------------|--------------|------|
| Adaptive prediction | 0            | 0    |
| Adaptive wavelet    | 2            | 0    |
| Wavelet transform   | 0            | 4    |

Table 2: Detection performance of each method.

In our detection scheme we use an adaptive threshold value which is determined from the first image of the video sequence according to the following formula:

$$T_h = \frac{A}{L} \sum_m h_m(I_1, I_2, I_3, I_4) \quad (6)$$

where  $h_m(I_1, I_2, I_3, I_4)$  represents the value of the statistic  $h$  inside the  $m^{th}$  block, and  $L$  represents the total number of blocks inside the image. In our case  $A$  is chosen as 12 so that  $T_h$  is well above the maximum  $h$  value of the blocks containing no moving target (maximum  $h$  value is 0.5 in the training set of 12 videos). Detection results are summarized in Table 2. In all of the 15 test videos the moving objects are determined successfully.

We also compared the performance of the adaptive predictor to the wavelet transform, and adaptive subband decomposition [7]. Motion compensated images are analyzed using (i) the adaptive predictor described in Section 2, (ii) wavelet transform (subband decomposition), and (iii) adaptive subband decomposition [7].

Typical results of the above methods are shown in Figure 4. The detection performance of these methods are summarized in Table 2. In the test videos, the adaptive predictor produces the best results. Adaptive subband decomposition also detects all of the moving objects but, in two cases, it produces false alarms. In ordinary wavelet transform, 4 targets are

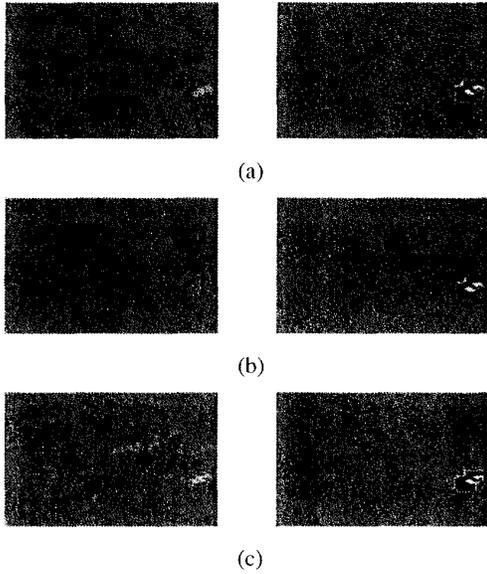


Figure 4: (a) Prediction error image  $e[m,n]$  and the detected region (right), (b) sum of the subimages  $x_{lh}[m,n]$  and  $x_{hl}[m,n]$  and the detected region (right), (c) sum of the subimages  $x_{lh}[m,n]$  and  $x_{hl}[m,n]$  obtained after adaptive subband decomposition and the detected region (right).

missed. By reducing the threshold all of the targets can be detected but in this case, the number of false alarms drastically increases.

## References

- [1] A. A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, and T. Sikora. "Image sequence analysis for emerging interactive multimedia services - the European COST 211 framework," *IEEE Trans. on CAS for Video Tech.*, Nov. 1998.
- [2] A. M. Tekalp, *Digital Video Processing*. Prentice-Hall, 1995.
- [3] R. Rajagopalan, E. Feig, and M. T. Orchard, "Motion optimization of ordered blocks for overlapped block motion compensation". *IEEE Trans. on CAS for Video Tech.*, Apr 1998.
- [4] M. N. Gürcan, Y. Yardımcı and A. E. Çetin, "Influence function based Gaussianity tests for detection of microcalcifications in mammogram images ", *ICIP'99*, 1999.
- [5] R. Ojeda, J. Cardoso, and E. Moulines, "Asymptotically Invariant Gaussianity Test for Causal invertible Time Series", *Proceedings of IEEE ICASSP'97*, vol. 5, pp 3713-3716, April 21-24, 1997.
- [6] G. B. Giannakis and M. K. Tsatsanis, "Time domain tests for Gaussianity and time-reversibility", *IEEE Trans. on Signal Processing*, vol. 42, pp. 3460-3472, 1994.
- [7] Ö. N. Gerek, A. E. Çetin, "Linear/nonlinear adaptive polyphase subband decomposition structure for image compression" *Proc. of IEEE ICASSP'98*, Seattle, U.S.A, May 1998.
- [8] O. Arıkan, A. E. Çetin, Engin Erzin, "Adaptive Filtering for non-Gaussian stable processes", *IEEE Signal Processing Letters*, vol. 1, No. 11, pages 163-165, Nov. 1994.