Signal Processing 82 (2002) 1941–1947

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Moving object detection using adaptive subband decomposition and fractional lower-order statistics in video sequences

A. Murat Bagci[a,1], Yasemin Yardimci[b], A. Enis Çetin[a,*]

[a] Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey

[b] Informatics Institute, Middle East Technical University, Ankara, Turkey

Received 1 May 2001; received in revised form 14 November 2001

Abstract

In this paper, a moving object detection method in video sequences is described. In the first step, the camera motion is eliminated using motion compensation. An adaptive subband decomposition structure is then used to analyze the motion compensated image. In the "low–high" and "high–low" subimages moving objects appear as outliers and they are detected using a statistical detection test based on fractional lower-order statistics. It turns out that the distribution of the subimage pixels is almost Gaussian in general. On the other hand, at the object boundaries the distribution of the pixels in the subimages deviates from Gaussianity due to the existence of outliers. By detecting the regions containing outliers the boundaries of the moving objects are estimated. Simulation examples are presented.

Keywords: Moving object detection; Adaptive subband decomposition; Wavelet transform; Fractional lower-order statistics

1. Introduction

In this paper, a moving object detection method in video sequences based on adaptive subband decomposition and fractional lower-order statistics (FLOS) is described. Detection of moving objects can be a complicated task especially when there is noise and the video camera is in motion. In some classical object detection methods [1,12,3,7], variances of the object and the background is compared to distinguish the object from the background. In this paper, we take advantage of the fact that objects produce outliers and local extrema in the motion compensated images and the wavelet (or subband) domain. We determine the object boundaries by detecting the regions having extrema and outliers using FLOS.

In our method, the first step is the elimination of the camera motion using motion compensation. After motion compensation, the resulting image basically contains the moving regions and objects. This image is further processed using a two-dimensional (2D) adaptive filter bank [5] in which the filters are updated according to a least mean square (LMS) type adaptation algorithm. In this filterbank structure, each pixel is adaptively predicted using an appropriate neighborhood structure and four subimages are obtained.

* Corresponding author.

E-mail address: cetin@ee.bilkent.edu.tr (A.E. Çetin).

[1] Now at Electrical and Computer Engineering Department, University of Illinois at Chicago, Chicago, IL, USA.

0165-1684/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0165-1684(02)00321-3

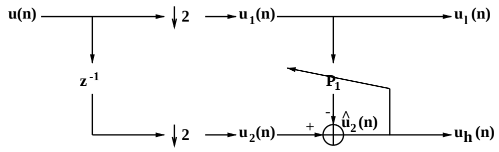

Fig. 1. Adaptive subband decomposition structure.

It turns out that the distribution of the "low–high" and "high–low" subimage pixels is almost Gaussian in general. However, moving objects produce outliers in the residual image as the pixels of the moving objects or their boundaries cannot be predicted accurately using the neighboring pixels. We detect the outliers using a fractional lower-order statistical test. In static regions the test statistic is close to zero whereas in regions containing the moving object(s) it produces high values. Subimages are analyzed in small blocks and moving objects are determined by estimating the FLOS-based statistic in each block.

In Section 2, we present the 2D adaptive subband decomposition method which tries to eliminate the static background in highbands. In Section 3, we review the FLOS-based statistical test that we use for moving object detection over highband subimages, and present the results of simulation studies in Section 4.

2. Adaptive subband decomposition

The concept of adaptive subband decomposition is developed in [4,5]. Adaptive subband decomposition can be considered as a trade-off between the adaptive prediction and ordinary lifting [11] based wavelet transform.

The adaptive subband decomposition structure [4–8] is shown in Fig. 1. The structure was developed for 1D signals, but we can apply it to 2D signals by using the row-by-row and column-by-column filtering methods as in 2D separable subband decomposition (or wavelet transform).

The first subsignal u_l is a downsampled version of the original signal u, a 1D signal which is usually a column or a row of the input image. As u_l is the result of a down-sampling by 2 operation, it contains only the even samples of the signal u. The sequence u_2 is a shifted and downsampled by 2 version of u, containing only odd samples of u. We predict u_2 using u_1 and subtract the estimate of u_1 from u_2 to obtain the signal u_h which contains unpredictable regions such as edges of the original signal.

Various adaptation schemes can be used for the predictor P_1 [5]. In our work, we used the adaptive FIR estimator, as it proved to be good for the sample images that have been tested. This adaptive FIR estimator is obtained by predicting the odd samples $u_2(n)$ from the even samples $u_1(n)$ as follows:

$$\hat{u}_2(n) = \sum_{k=-N}^{N} w_{n,k} u_1(n-k) \tag{1}$$

or

$$\hat{u}_2(n) = \sum_{k=-N}^{N} w_{n,k} u(2n-2k). \tag{2}$$

The filter coefficients $w_{n,k}$'s are updated using an LMS-type algorithm as follows:

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu \frac{\tilde{\mathbf{v}}_n e(n)}{\|\tilde{\mathbf{v}}_n\|^2}, \tag{3}$$

where $\hat{\mathbf{w}}(n) = [w_{n,-N}, \ldots, w_{n,N}]$ is the weight vector at time instant n,

$$\tilde{\mathbf{v}}_n = [u_1(n-N), u_1(n-N+1), \ldots, u_1(n+N)]^{\mathrm{T}}. \tag{4}$$

The subsignal u_h is given by

$$u_h(n) = u_2(n) - \hat{u}_2(n), \tag{5}$$

where u_h is the error we make in predicting the odd samples from the even samples, thus,

$$e(n) = u_h(n) = u_2(n) - \tilde{\mathbf{v}}_n^{\mathrm{T}}(n)\hat{\mathbf{w}}(n). \tag{6}$$

Both ℓ^1 and ℓ^2 norms can be used in normalizing the update equation in (3) depending on the characteristics of the signal [2]. The use of ℓ^1 norm in (3) produces more robust results, if the images are corrupted by salt and pepper type noise which can be modelled via α-*stable* random process or epsilon contaminated Gaussian process concept. In this paper, the images are directly obtained from either a CCD camera or an infrared camera and they are almost noise free. Therefore, regular Euclidian norm is used in experimental studies. For the initial filter one can use a typical lowpass FIR filter of length $2N+1$ for the adaptive predictor. The convergence of the adaptive filter is observed to be fast in natural images, and we have not observed any divergence problem in all images that we have analyzed.

This structure is the simplest adaptive filterbank. Other adaptive filterbanks in which the "low-band" subsignal is a lowpass filtered and downsampled version of the original signal can be found in [5].

If the motion compensated image is processed by an adaptive filterbank we expect that small moving object boundaries cannot be predicted as good as the other static pixels. Thus outliers and/or local extrema will appear in $u_h[n]$ in regions corresponding to moving objects.

The extension of the adaptive filterbank structure to two dimensions is straightforward. As in the case of ordinary subband decomposition, we process the image rowwise first and obtain two subimages. Consequently, these two subimages are processed columnwise and the low–low subimage x_{ll}, the low–high subimage x_{lh}, the high–low subimage x_{hl}, and the high–high subimage x_{hh} are obtained.

In general, the "low–high" and "high–low" images are sharper (smother) at the edges of the objects (static image regions) in adaptive subband compared to regular subband decomposition. This is due to the fact that static pixels can be predicted very effectively using the neighboring pixels whereas the pixels belonging to moving objects cannot be predicted from the background pixels. Adaptive subband decomposition gives better results in moving target detection for this reason.

3. Fractional lower-order statistical test

In our approach, the video containing a moving object(s) is (are) analyzed as follows:

- A motion compensated image is obtained from two consecutive images [10].
- Adaptive subband decomposition of the motion compensated image is computed.
- The resulting subimages $x_{lh}[m,n]$ and $x_{hl}[m,n]$ are summed and analyzed block-by-block by using the lower-order statistical detection test, and
- The blocks in which the lower-order statistics exceeds a threshold are marked as the region(s) containing the moving object.

In Fig. 2, an image of a moving minivan extracted from a video is shown. The motion compensated image obtained using this image frame and the next one is shown in Fig. 3. In this video the camera is fixed, therefore, the image shown in Fig. 3 is simply obtained by subtracting the two consecutive image frames from each other. In Fig. 4, the subimage x_{lh}, and in Fig. 5 the subimage x_{hl} are shown, respectively.

Fig. 2. An image of a moving minivan from a video sequence.

Fig. 3. Motion compensated image.

It is experimentally observed that in regions with no moving objects, the subimages $x_{lh}[m,n]$ and $x_{hl}[m,n]$ have Gaussian like distribution in most natural images whereas regions containing moving objects have outliers and the distribution of pixels deviate from Gaussianity (the high–high subimage $x_{hh}[m,n]$ contains almost no information for most practical images and it is not used in our algorithm). The appearance of outliers at object boundaries in subimages is due to the fact that pixels of a moving object cannot be accurately predicted using the surrounding pixels as shown in Figs. 4 and 5.

Fig. 4. The low–high subimage obtained using adaptive subband decomposition of the motion compensated image.

Fig. 5. The high–low subimage obtained using adaptive subband decomposition of the motion compensated image.

In [3,7], variance or power is used to distinguish the objects from the background in the motion compensated image. It is assumed that the object and the background have different variances. Since the data that we analyze is essentially non-Gaussian and contains outliers due to moving objects FLOS is used instead of variance in this paper. The use of FLOS brings robustness and reduces the number of false alarms.

Recently Gonzales and Arce [6] proposed a framework called zero-order statistics to analyze very impulsive processes, and they defined a statistic called geometric power. We use the geometric power as a test statistic in the analysis of motion compensated image. The geometric power is defined as

$$\hat{S}_0 = \exp\left(\frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} \log |e[m,n]| \right), \tag{7}$$

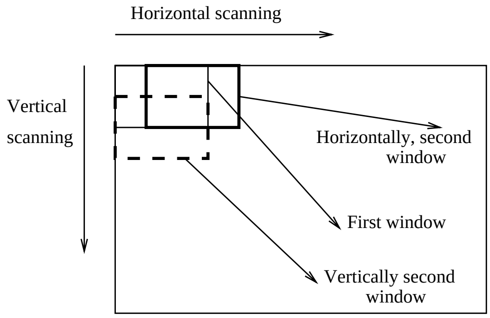

Fig. 6. Computation of the test statistic in overlapping windows.

where $e[m,n]$ represents the sum of the pixel values $x_{lh}[m,n]$ and $x_{hl}[m,n]$ and $M \times N$ is the size of the region in which $\hat{S}_0$ is estimated. As pointed above, the subimages x_{lh} and x_{hl} are obtained by processing the motion compensated image using the adaptive subband decomposition. The high–high subimage $x_{hh}[m,n]$ contains almost no information for most practical images and it may contain noise thus it is not used in our algorithm. The statistic $\hat{S}_0$ can also be expressed as follows:

$$\hat{S}_0 = \left(\prod_{m=1}^{M} \prod_{n=1}^{N} |e[m,n]| \right)^{1/(M \times N)}. \tag{8}$$

Subband images, x_{lh} and x_{hl}, are zero-mean images as they do not contain any low-frequency information (Figs. 4 and 5). In static regions pixels of x_{lh} and x_{hl} are close to zero. Therefore, we expect that the geometric power takes small values in static image regions and it should take large values around moving objects due to outliers in $e[m,n]$.

We divide the image to be analyzed into M by N blocks. The FLOS-based statistic (8) is calculated within each block inside the image. These blocks may overlap as shown in Fig. 6. In our experimental work we used blocks of size $M = 8$ by $N = 8$ where overlapping occurs at 4 pixel steps. If the FLOS-based statistic exceeds a threshold value in a block then this block is marked as a region containing a moving object or part of a moving object if the object size is larger than 8×8. The above procedure is carried out over the entire video sequence.

As described above in each image block a statistical test is carried out to detect the moving object(s).

The detection procedure 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: $\hat{S}_0 < T_h$,
- H_1: $\hat{S}_0 \geqslant T_h$.

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

Another statistical detection approach is based on estimating the parameter α of Symmetric α-stable distribution in overlapping image blocks. We expect that the parameter α should be close to 2 in static regions where the distribution of image $e[m,n]$ pixels is almost Gaussian, and α takes lower values than 2 around moving objects due to outliers in $e[m,n]$.

4. Experimental results

In this section, we present simulation studies. We test the performance of the detection scheme by analyzing 10 video sequences containing moving objects on various backgrounds. As described in Sections 1 and 2, motion compensated images are obtained in the first step. A classical block matching based motion compensation algorithm with subpixel accuracy is used [12].

In the second step, motion compensated images are filtered using the adaptive wavelet transformer and the subimages $x_{lh}[m,n]$ and $x_{hl}[m,n]$ are obtained. Finally, the test statistic values are obtained in small overlapping blocks.

In our detection scheme we use an adaptive threshold value which is determined from the first two images of the video sequence. The image $e[m,n]$ is divided into three horizontal strips. In each strip the mean and the variance of the test statistic is estimated and a threshold is determined for each strip as follows:

$$T_{h,i} = \mu_i + \lambda\sigma_i, \quad i = 1,2,3, \tag{9}$$

where μ_i and σ_i are the mean and the standard deviation of the test statistic in the strip i, respectively. The parameter λ is usually selected as 3 as a rule of thumb which is based on the fact that in regular distributions including the Gaussian distribution almost all of the observations fall within the segment determined by the 3 σ_i. Anything exceeding the threshold $T_{h,i}$ is considered to be an outlier. In our experiments the parameter λ is selected as 2.5 to further reduce the rate of missed targets.

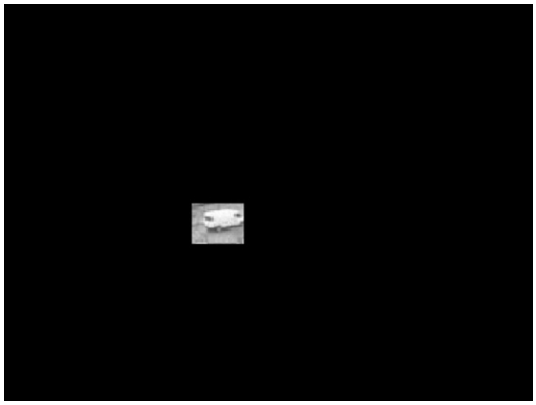

Fig. 7. The detected moving object: Regions exceeding the threshold based on FLO statistic, geometric power.

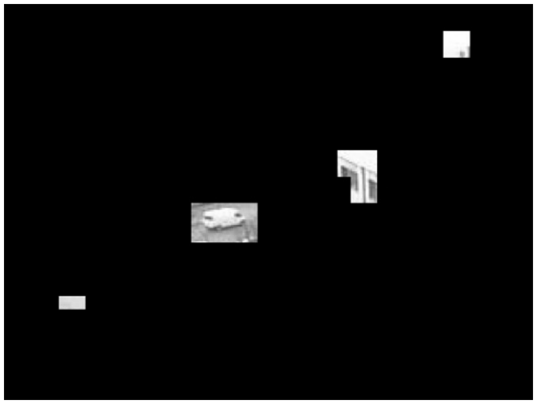

Fig. 8. Regions exceeding the variance-based threshold.

The image shown in Fig. 7 shows the small regions exceeding the threshold based on the geometric power, the FLO statistic defined in Eq. (7). The minivan shown in Fig. 2 is clearly detected. The image shown in Fig. 8 shows the small regions exceeding the variance-based threshold. The minivan is detected but there are four other false alarms.

In all of the 10 test videos the moving targets are successfully detected. 26 detection results are summarized in Table 1. In these detection experiments

Table 1
Comparison of variance, geometric power and HOS-based detection methods in 26 different scenarios in 10 videos

Detection method		Variance based		HOS based		Geometric power based	
Scenario	Number of targets in frame	False alarms	Miss	False alarms	Miss	False alarms	Miss
1	1	0	0	1	0	2	0
2	1	6	0	0	0	1	0
3	1	1	0	0	0	0	0
4	1	1	0	2	0	2	0
5	1	4	0	6	0	4	0
6	1	4	0	0	0	2	0
7	1	2	0	0	0	3	0
8	2	2	0	0	0	1	0
9	2	4	0	1	0	1	0
10	1	9	0	0	0	0	0
11	1	7	0	0	0	0	0
12	2	3	0	0	0	0	0
13	3	4	0	1	0	2	0
14	4	5	0	2	0	1	0
15	4	2	0	1	0	1	0
16	1	2	0	0	0	2	0
17	1	4	0	0	0	1	0
18	1	4	0	0	0	1	0
19	3	2	0	1	1	0	0
20	3	2	0	0	0	0	0
21	3	0	0	0	0	0	0
22	1	2	0	0	0	2	0
23	1	7	0	7	0	4	0
24	2	1	1	0	1	0	1
25	1	3	0	9	1	2	0
26	1	3	0	4	0	2	0
Total	43	84	1	35	3	34	1

the number of false alarms for variance, higher-order statistics (HOS) and geometric power-based detection methods are 3.23, 1.35, and 1.31 per image, respectively. The use of geometric power significantly reduces the number of false alarms compared to the variance-based detection method. Miss rate of geometric power-based method is less than the HOS-based test statistic which utilizes third- and fourth-order correlations [13].

Variance or geometric power-based detection methods rarely miss moving objects in all the videos that we have tried. Even if a moving object is missed in the current and previous image frames it is always detected in the next two or three image frames.

The performance of the adaptive predictor to the wavelet transform, and adaptive subband decomposition [5] is compared in [13]. If regular subband decomposition is used instead of adaptive subband decompositions then in the above data set the false alarm rates increase to 8.65 per image for variance-based detection method and 1.92 per image in FLOS-based detection method, respectively. In general, adaptive subband decomposition provides a good trade-off between regular 2D adaptive prediction and the ordinary wavelet transform in terms of detection performance and the computational cost.

The computational cost of the adaptive prediction-based method [13] is much higher than the adaptive subband decomposition-based method in which a quarter size image $x_{lh} + x_{hl}$ is analyzed. Whereas in adaptive prediction-based method FLOS test computations are carried out over the entire image x.

5. Conclusion

In this paper, a moving target detection method is proposed. The method is based on adaptive subband decomposition and fractional lower-order statistics. Experimental results indicate that the proposed method is more robust compared to second-order statistics based methods.

The FLOS-based detection method can be combined with other segmentation clues as in [7] to achieve an automatic detection of the moving objects from the background.

The new video coding standard MPEG-4 [1,7] is an object-based method in the sense that objects in video can be defined and coded separately. Due to this reason the problem of object boundary estimation receive a lot of attention [3,7,9]. The proposed FLOS-based method can be used for this application as well. In our approach a tight region containing the moving object is determined. Detecting the exact boundary of the object within this region is a much easier problem than analyzing the entire image. For example, the active-contour-based boundary detection method proposed in [9] can be applied inside the detected region instead of the entire frame.

The proposed method is computationally efficient as the detection operation is carried out over quarter size subband images instead of the full size image frame.

References

[1] A.A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, T. Sikora, Image sequence analysis for emerging interactive multimedia services—the European COST 211 framework, IEEE Trans. Circuits and Systems Video Technol. 8 (7) (November 1998) 802–813.

[2] O. Arikan, A.E. Cetin, E. Erzin, Adaptive filtering for non-Gaussian stable processes, IEEE Signal Process. Lett. 1 (11) (November 1994) 163–165.

[3] A. Ekin, A.M. Tekalp, R. Mehrotra, Automatic extraction of low-level object motion descriptors, Proceedings of the International Conference on Image Processing, Thessaloniki, Vol. 2, October 2001, pp. 633–636.

[4] O.N. Gerek, A.E. Cetin, Polyphase adaptive filter banks for fingerprint image compression, Electron. Lett. 34 (20) (October 1998) 1931–1932.

[5] O.N. Gerek, A.E. Cetin, Adaptive polyphase subband decomposition structures for image compression, IEEE Trans. Image Process. 9 (10) (October 2000) 1649–1660.

[6] J. Gonzales, G.R. Arce, Zero-order statistics: a signal processing framework for very impulsive processes, Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics, Banff, Canada, July 1997, pp. 254 –258.

[7] M. Kim, J.C. Choi, D. Kim, H. Lee, C. Ahn, Y-S. Ho, A VOP generation tool: automatic segmentation of moving objects in image sequences based on spatio-temporal information, IEEE Trans. Circuits and Systems Video Technol. 9 (8) (December 1999) 1216–1226.

[8] R. Oktem, K. Egiazarian, A.E. Cetin, Subband decomposition based image compression algorithms with nonlinear adaptive filter banks, Proceedings of the IEEE- EURASIP NSIP 99, Antalya, Turkey, Vol. 2, June 1999, pp. 766–769.

[9] F. Preceiso, M. Barlaud, B-spline active contours for fast video segmentation, Proceedings of the International Conference on Image Processing, Thessaloniki, Vol. 2, October 2001, pp. 777–780.

[10] R. Rajagopalan, E. Feig, M.T. Orchard, Motion optimization of ordered blocks for overlapped block motion compensation, IEEE Trans. Circuits and Systems Video Technol. 8 (2) (April 1998) 119–123.

[11] W. Sweldens, The lifting scheme: a new philosophy in biorthogonal wavelet constructions, Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE), Vol. 2569, September 1995, pp. 68–79.

[12] A.M. Tekalp, Digital Video Processing, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[13] R. Zaibi, Y. Yardimci, A.E. Cetin, Small moving object detection in video sequences, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2000), Istanbul, Turkey, Vol. 4, June 2000, pp. 2071–2074.