MOVING OBJECT DETECTION IN VIDEO BY DETECTING NON-GAUSSIAN REGIONS IN SUBBANDS AND ACTIVE CONTOURS

M.Yagmur Gok[1], A. Enis Cetin[2]

[1]MBDF, Sabanci University, Istanbul, Turkey
[2]Department of Electrical and Electronics Engineering, Bilkent University TR-06553,Ankara,Turkey

ABSTRACT

A multi-stage moving object detection algorithm in video is described in this paper. First, the camera motion is eliminated by motion compensation. An adaptive subband decomposition structure is then used to analyze the difference image. In the high-band subimages, moving objects which produce outliers are detected using a statistical test determining non-Gaussian regions. It turns out that the distribution of the subimage pixels is almost Gaussian in general. But, at the object boundaries the distribution of the pixels in the subimages deviates from Gaussianity due to the existence of outliers. Regions containing moving objects in the original image frame are detected by detecting regions containing outliers in subimages. Finally, active contours are initiated in these regions in the wavelet domain and object boundaries are accurately estimated.

1. Introduction

In this paper,a multistage moving object detection method based on subband analysis of the video is described. In the first stage of our algorithm both temporal differencing and-or background subtraction methods [1]-[2] can be used after compensating the camera motion using a standard method. In the second stage of our algorithm, we perform wavelet analysis to robustify the differencing operation. In addition, wavelet analysis allows us to use a statistical non-Gaussian region detection method to determine regions containing moving objects. Our Algorithm is based on the fact that moving objects produce outliers and local extrema in the wavelet (or subband) domain . It is observed by many researchers that the distribution of wavelet coefficients tends to be Gaussian in natural images. On the other hand, the distribution deviates from the normal distribution around moving objects in the wavelet domain. To determine the regions containing outliers we studied two non-Gaussian region detection tests based on Higher Order Statistics (HOS) and Fractional Lower Order Statistics (FLOS). In some classical object detection methods [3], variances of the object and the background is compared to distinguish the object from the background. We compared the FLOS, HOS and variance based statistical tests experimentally and observed that the FLOS based test produced the best results. In the

third stage of the algorithm the wavelet subimages are divided into small overlapping blocks and in each block FLOS based statistic is estimated. In static regions the test statistic is close to zero whereas in regions containing the moving object(s) it produces high values. Regions containing moving objects are determined by thresholding the test statistic in small blocks. In the last stage of the algorithm active contours are initiated in the detected regions to estimate the object boundaries.

In Section 2, we present the 2-D adaptive wavelet analysis method which reduces the effects of the static background in highband subbands. In Section 3, we review the FLOS based statistical tests that we use for moving object detection in highband subimages. In Setion 4, we use active contours to determine the object boundaries, and present the results of simulation studies in Section 5.

2.Adaptive Wavelet Transform

In most moving object detection methods, frame differencing or background subtraction is used to highlight moving regions. Boundaries of the static objects are also emphasized by differencing due to undersampling of images. In addition noise may be amplified. To robustify the differencing operation we use adaptive wavelet analysis and remove the highest frequency subband which may correspond to amplified high frequency noise components. Adaptive wavelet filterbank removes the correlation between the neighboring image samples using a Least Mean Square (LMS) type adaptation strategy. In this way, we suppress the static background pixels which are highly correlated with each other and highlight the moving objects from the background as moving object pixels are uncorrelated with the background pixels. The concept of adaptive (wavelet) subband decomposition is developed in [4]. Adaptive subband decomposition structure is obtained by using adaptive LMS filters instead of fixed filters in the lifting based wavelet transform. One dimensional adaptive subband decomposition structure [6,8,9] is shown in Figure 1. Its extension to two-dimensional (2-D) signals is straightforward by using the row by row and column by column filtering methods as in ordinary 2-D separable subband decomposition (or wavelet transform). The first subsignal v_1 is a downsampled version

of the original signal v, a one dimensional signal which is usually a column or a row of the input image. As v_1 is obtained after down-sampling, it contains only the even samples of v. The sequence v_2 is a shifted and downsampled version of v, containing only odd samples of v. We predict v_2 using v_1 and subtract the estimate, $\hat{v}_2$ from v_2 to obtain the signal v_h which contains unpredictable regions such as edges of the original signal. Various adaptation schemes can be used for the predictor P_1 [4]. In our work, we used the adaptive FIR estimator, proven to be good for the sample images that have been tested. This adaptive estimator is obtained by predicting the odd samples $v_2(n)$ from the even samples $v_1(n)$) using an LMS-type algorithm as follows:

$$\widehat{v}_2(n) = \sum_{k=-N}^{N} \omega_{n,k} v_1(n-k) = \sum_{k=-N}^{N} \omega_{n,k} v(2n-2k)$$

(1)

Filter coefficients $\omega_{n,k}$ are updated using an LMS type algorithm as follows:

$$\widehat{w}(n+1) = \widehat{w}(n) + \mu \frac{\overline{v}_n e(n)}{\|\overline{v}_n\|^2}.$$

(2)

where $\widehat{w}(n) = [\omega_{n,-N},...,\omega_{n,N}]$ is the weight vector at time instant n, $e(n) = v_h(n) = v_2(n) - \overline{v}^T(n)\overline{w}(n)$, and $\overline{v}_n = [v_1(n-N), v_1(n-N+1),..., v_1(n+N-1), v_1(n+N)]^T$. The highband subsignal v_h is given by

$$v_h(n) = v_2(n) - \widehat{v}_2(n).$$

(3)

This structure is the simplest adaptive filterbank. Other adaptive filterbanks in which the "low-band" subsignal is a low-pass filtered and downsampled version of the original signal can be found in [4]. If the differenced image is processed by an adaptive filterbank we expect that moving object boundaries are not predicted as good as the other static pixels. Thus outliers and/or local extrema will appear in $v_h[n]$ in regions corresponding to moving objects.

The extension of the adaptive filterbank structure to two dimensions is straightforward. We first process the image x row-wise then columnwise, and obtain four subimages, x_{ll}, x_{lh}, x_{hl}, and x_{hh}. In general, the 'low-high' and 'high-low' images obtained are sharper (smother) at the edges of the objects (static image regions) in adaptive subband decomposition compared to the regular subband decomposition. This is due to the fact that static pixels can be predicted effectively using the neighboring pixels whereas the pixels belonging to moving objects can not.

3. Detection of Moving Regions in Subbands

In our approach, a region containing the moving object(s) is(are) determined as follows: 1.A motion compensated difference image is obtained from two or three consecutive image frames of the video (background subtraction can be also used, if the camera is stationary), 2. adaptive wavelet analysis of the motion compensated difference image is carried out, 3. the resulting subimages $x_{lh}[m,n]$ and $x_{hl}[m,n]$ are summed and analyzed block by block by using a non-Gaussian region detection test, and 4.the blocks in which the test statistic exceeds a threshold are marked as the region(s) containing (portions of) the moving object.

Recently, Gonzales and Arce [5] proposed a Fractional Lower Order Statistics (FLOS) to analyze impulsive random processes, and they defined a statistic called Geometric Power (GP). In moving object detection we use the geometric power as a test statistic in the analysis of motion compensated difference image. The geometric power is defined as

$$\widehat{S}_o = \exp(\frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} \log|e[m,n]|).$$

(4)

where $e[m,n]$ represents the sum of the pixel values $x_{lh}[m,n]$ and $x_{hl}[m,n]$ and $M \text{x} N$ is the size of the region in which S_o is estimated. As pointed above the subimages, x_{lh} and x_{lh} are obtained by processing the motion compensated difference image using the adaptive subband decomposition. The high-high subimage $x_{hh}[m,n]$ contains almost no information. The statistic S_o can also be expressed as follows: $\widehat{S}_o = (\prod_{m=1}^{M} \prod_{n=1}^{N} \log|e[m,n]|)$. Subband images, x_{lh} and x_{hl}, are zero-mean images as they do not contain any low frequency information. 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 $e[m,n]$. As discussed in the next section it is experimentally observed that FLOS based statistic produces better results than the HOS based statistic.

We divide the image to be analyzed into M by N blocks. The FLOS based statistic (9) is calculated within each block. These blocks may overlap. 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 by 8.

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 pixels within a block of data:

$$H_0 : \widehat{S}_0 < T_h$$

$$H_1 : \widehat{S}_0 \geq T_h$$

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.

In [3], [6] variance or power is used to distinguish the objects from the background which has a different variance.

Variance based detection produces a lot of false alarms in the videos that we have tried compared to FLOS and HOS based test statistics. It turns out that the Fractional Lower Order Statistic (FLOS) produces the best results. The use of FLOS brings robustness and reduces the number of false alarms. This approach cannot be used to detect moving point orvery small targets which are treated as noise by both the HOS and FLOS based test statistics.

5. Active Contours

The last step of our moving object detection algorithm is to determine the boundaries of the moving objects. The problem of object boundary estimation receive a lot of attention [3,6,8,9]. Active contours can be used for this purpose. We carry out boundary detection in the wavelet domain using the highband subimages x_{hl} and x_{lh}. The 2-D signal $|x_{hl}| + |x_{lh}|$ contains both horizontal and vertical edges of the original object. Therefore an active contour initiated from the edges of the rectangular region determined by the statistical test converges to the boundary of the object. We place the initial snaxels around the edges of the region detected by the FLOS test which contains the moving object. Snake contour estimation algorithms are greedy algorithms minimizing a linear combination of the so-called image energy, curvature energy, and the continuity energy. They may converge to local minima. In our case, the minimization is carried out in the wavelet domain based on the pixel values of $|x_{hl}| + |x_{lh}|$. In all the convex examples that we tried they converge to the boundary of the objects.

In this paper, decimated wavelet transform whose space resolution is lower than actual data is used. If higher accuracy is desired then undecimated wavelet transform may be used to determine the boundary of a moving object, or the last few iterations of the contour algorithm can be implemented over the actual image.

In Figure 10, the final control points of a snake curve is shown over a helicopter image from an infrared test video.

4. Experimental Results

The performance of the detection scheme is tested by analyzing 10 infrared FLIR videos and two regular videos containing moving objects on various backgrounds.

In the first step, a classical block matching based motion compensation algorithm with subpixel accuracy is used. 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 adaptive threshold values which are determined from the first two images of the video sequence and updated after several frames. The image $e[m, n] = |x_{lh}[m, n]| + |x_{hl}[m, n]|$ 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

$$T_{h,i} = \mu_i + \lambda \sigma_i \qquad (5)$$

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 $3\sigma_i$. Any block in which the test statistic exceeds the threshold T_h is considered to contain outliers or equivalently a portion of the moving object. In our experiments the parameter λ is selected as 2.5 to further reduce the rate of missed targets. In Figure 2, an image of a moving minivan extracted from a video is shown. The images in Figure 3 show the union of the small regions exceeding the threshold based on variance, and the geometric power, the FLO statistic defined in Equation (4). The minivan shown in Figure 2 is detected by both methods. If we carry out hypothesis testing based on variance of the wavelet coefficients the minivan is detected but there are four other false alarms. The output of the FLOS based system is also shown in Figure 3 in which there are no false regions. In all of the 10 test infrared videos and two regular the moving targets are successfully detected. In these detection experiments the number of false alarms for variance, Higher Order Statistics (HOS) and FLOS 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 second order correlations [7]. The HOS or geometric power (FLOS) 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 image frame it is always detected in the next two or three image frames.

In Figure 4, snaxels of the active contour around the minivan is shown. 12 snaxels characterizing the active contour are initiated in the wavelet domain from the edges of the detected region shown in Figure 3. In Figure 5, a helicopter from an infrared FLIR video is shown. Final snaxel locations are also marked on this image.

6. Conclusion

In this paper, a moving object detection algorithm is proposed. The method is based on adaptive subband decomposition and non-Gaussian region detection in subband images. Experimental results indicate that the proposed method is more robust compared to the first and second order statistics based methods in which the difference image data is thresholded for detection.

Non-Gaussian regions in subbands corresponding to moving objects are detected using a FLOS test which can be easily combined with other segmentation clues to achieve more complex moving object detection systems. The pro-

posed method is computationally efficient as the detection operation is carried out over quarter size subband images instead of the full size image frame. By using an ordinary fixed wavelet filterbank the computational cost can be further reduced as there are many computationally very efficient (Order(N)) wavelet transforms. In the last step of the algorithm active contours are initiated in detected regions in the wavelet domain and object boundaries are accurately estimated.

References

[1] Takeo Kanade, Robert T. Collins, Alan J. Lipton, et al., in Cooperative Multi-Sensor Video Surveillance," in Proceedings of DARPA Image Understanding Workshop, volume 1, pages 324, 1998.

[2] I.Haritaoglu, D.Harwood, and L.Davis, W4: Who, When, Where, What: A Real Time System for Detecting and Tracking People," Third Face and Gesture Recognition Conference. pages:222-227 1998

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

[4] Omer N. Gerek, A. Enis Cetin, "Adaptive Polyphase Subband Decomposition Structures for Image Compression," *IEEE Trans. on Image Proc.,* pp. 1649-1660, Oct. 2000.

[5] J. Gonzales, G. R. Arce, "Zero-Order Statistics: A signal processing framework for very impulsive processes", Proc. of the IEEE Signal Processing Workshop on Higher-Order Statistics, Banf, Canada, July 1997, pp. 254-258.

[6] 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 Transactions on Circuits and Systems for Video Technology, Vol. 9 Issue: 8 , Dec. 1999, pp. 1216 -1226,

[7] R. Zaibi, Y. Yardimci, A. E. Cetin, "Small Moving Object Detection In Video Sequences", Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2000), Istanbul, Turkey, Vol4. June 2000, pp. 2071-2074.

[8] F. Preceiso, M. Barlaud, B-spline active contours for fast video segmentation," Proceedings International Conference on Image Processing, Thessaloniki, Oct 2001, Vol. 2, pp. 777-780.

[9] A. E. Cetin, et. al "Identification of relative protein bands in on Polyacrylamide Gel Electrophoresis(PAGE) images using multi-resolution snake algorithm," *Biotechniques*, 1162-1169, June 1999.
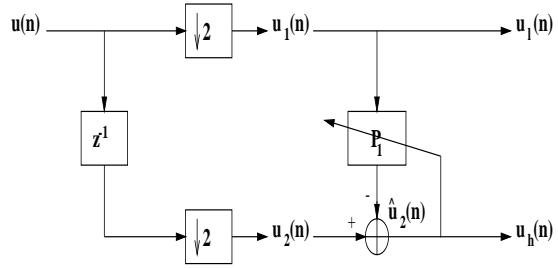
Figure 1: One-dimensional adaptive subband decomposition structure.

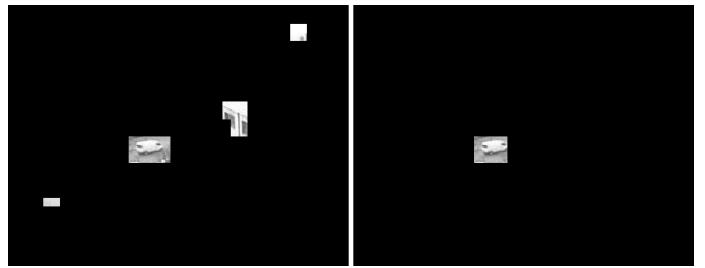Figure 2: A moving minivan image from a video.

Figure 3: Detected regions using the variance (left) and FLOS (right) based methods.

Figure 4: Estimated snaxels of the active contour are dark points on the edges of the minivan.

Figure 5: A helicopter image from an infrared video. Estimated snaxels of the active contour are dark points on the edges of the helicopter.