

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

Small moving object detection using adaptive subband decomposition and fractional lower order statistics in video sequences

A. Murat Bagci
Yasemin C. Yardimci
Enis A. Cetin

SPIE.

Small Moving Object Detection Using Adaptive Subband Decomposition and Fractional Lower Order Statistics in Video Sequences

A. Murat Bagci*, Yasemin Yardimci[†] and A. Enis Çetin**

**Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey*

[†]*Informatics Institute, Middle East Technical University, Ankara, Turkey*

***Sabanci University, Istanbul, Turkey*

Abstract. In this paper, a small moving object method 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 highband subimages moving objects appear as outliers and they are detected using a statistical detection test based on lower 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. 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, lower order statistics.

INTRODUCTION

In this paper, a moving object method detection method in video sequences based on adaptive subband decomposition and lower order statistics 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],[2],[3],[4] 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 containing outliers.

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 further is processed using a two-dimensional (2-D) adaptive filter bank [5] in which the filters are updated according to an 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. 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 lower order statistical test. In static regions the test statistics is closer to zero compared to regions containing the moving object(s).

In Section 2, we present the 2-D adaptive subband decomposition method which removes the static background. In Section 3, we review the Fractional Lower Order Statistics (FLOS) based test that we used for moving object detection and present the results of simulation studies in Section 4.

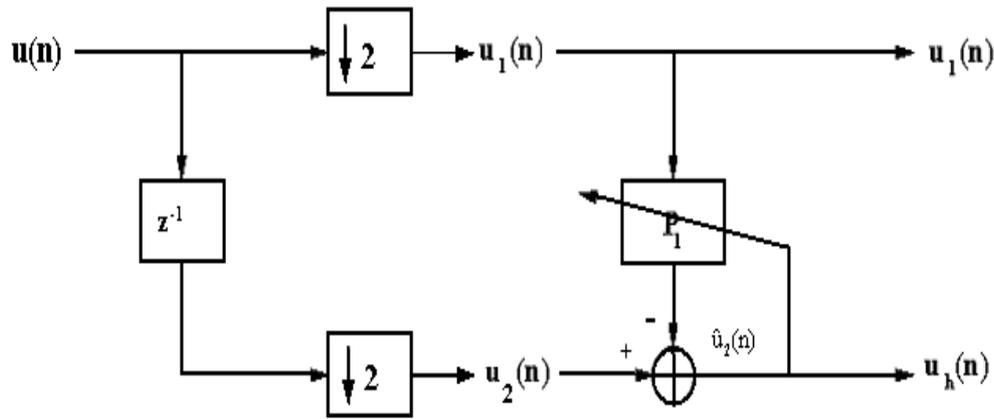


FIGURE 1. Adaptive subband decomposition structure.

ADAPTIVE SUBBAND DECOMPOSITION

The concept of adaptive subband decomposition is developed in [5, 6]. 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 [5]-[8] is shown in Figure 1. The structure was developed for one-dimensional signals, but we can apply it to two-dimensional signals by using the row by row and column by column filtering methods as in 2-D separable subband decomposition (or wavelet transform).

The first subsignal u_1 is a downsampled version of the original signal u , a one dimensional signal which is usually a column or a row of the input image. As u_1 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 . 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:

$$\hat{u}_2(n) = \sum_{k=-N}^N w_{n,k} u_1(n-k) = \sum_{k=-N}^N w_{n,k} u(2n-2k) \quad (1)$$

The filter coefficients $w_{n,k}$'s are updated using an LMS-type algorithm [12] as follows:

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

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

$$\tilde{\mathbf{v}}_n = [u_1(n-N), u_1(n-N+1), \dots, u_1(n+N-1), u_1(n+N)]^T, \quad (3)$$

The subsignal u_h is given by

$$u_h(n) = u_2(n) - \hat{u}_2(n). \quad (4)$$

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

$$e(n) = u_h(n) = u_2(n) - \tilde{\mathbf{v}}_n^T(n) \hat{\mathbf{w}}(n). \quad (5)$$

Both ℓ^1 and ℓ^2 norms can be used in normalizing the update equation in (2) depending on the characteristics of the signal [12]. In this paper, the regular Euclidian norm is used. For the initial filter one can use a typical lowpass filter for the adaptive predictor. The convergence of the adaptive filter is observed to be fast in natural images.



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



FIGURE 3. Motion compensated image.

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 four subimages x_{ll} , x_{lh} , x_{hl} , and x_{hh} are obtained. Figure 2 shows the original image x , and the resulting subimages x_{ll} , x_{lh} , x_{hl} and x_{hh} obtained after adaptive subband decomposition are shown in Figure 4, respectively.



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



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

This image is also processed by an ordinary wavelet transform. The resulting subimages are shown in Figure 10. The ‘low-high’ and ‘high-low’ images are sharper and the edges of the objects are highlighted more in the adaptive subband case. Adaptive subband decomposition gives better results in moving target detection for this reason.

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,
- 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 detection threshold are exceeded are marked as the region(s) containing the moving object.

In Figure 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 Figure 3. In this video the camera is fixed therefore the image shown in Figure 3, is simply obtained by subtracting the two consecutive image frames from each other. In Figure 4, the subimage x_{lh} , and in Figure 5 the subimage x_{hl} are shown, respectively.

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 contain 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 Figures 4 and 5.

In [3],[4] 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 Fractional Lower Order Statistic (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 [10] proposed a framework called zero order statistics to analyze very impulsive processes, and they defined a statistic called geometric power. We use the the geometric power as a test statistic in the analysis of motion compensated image. The geometric power is defined as

$$\hat{S}_o = \exp\left(\frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N \log |e[m, n]|\right) \quad (6)$$

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}_o 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}_o can also be expressed as follows:

$$\hat{S}_o = (\prod_{m=1}^M \prod_{n=1}^N |e[m, n]|)^{\frac{1}{M \times N}} \quad (7)$$

We expect that the geometric power takes small values around static image regions and it should take large values around moving regions due to outliers in $e[m, n]$.

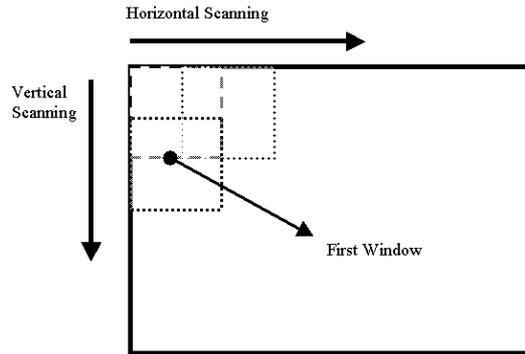


FIGURE 6. Computation of the test statistic in overlapping windows.

We divide the image to be analyzed into M by N blocks. The FLOS based statistic (7) is calculated within each block inside the image. These blocks may overlap as shown in Figure 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 by 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:

- $H_0: \hat{S}_o < T_h$
- $H_1: \hat{S}_o \geq T_h$

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 the small 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 pixels is Gaussian in $e[m, n]$ and α takes lower values than 2 around moving objects due to outliers in $e[m, n]$.

DETECTION OF POINT TARGETS IN IR IMAGES

In this section we describe the methods we use for detection of point (one pixel) targets, with sub-pixel velocity per frame. The images we use are obtained using an airborne infra-red camera. The resolution is 12 bits/pixel, but the data is limited in bandwidth so the image is stretched linearly to fit in the range of 0-255. All of the images are in a 320 column by 244 row format.

A sample frame from an infra-red video sequence is shown in Figure 7. The target is shown inside the rectangle. The original image does not contain the rectangle. The image sequences are obtained from the Air Force Research Laboratory web site ¹.

We make tests with five different detection methods. The results are included in Table 1. The numbers under each method denote the rank given to the target assigned by that method. A sample ranked image is given in Figure 8. The tests are performed on two sets of sequences.

The methods we propose are as follows. We first eliminate the background clutter using a motion compensation operation. The videos include targets with sub-pixel velocity per frame so we perform differencing between frames having a gap of 0.2 seconds. This corresponds to a difference of 5 image frames. In the first sequence the target is fast compared to the target in the second sequence. This explains why the detection methods perform better in the first sequence compared to the second.

Two methods based on adaptive wavelet transform are studied. In these methods the difference image is filtered with adaptive wavelet transform. The second step is extraction of the local maxima in the residual image. In third step the residual image is divided into 3 by 3 blocks and either a fourth order test or FLO statistics are calculated around the local maxima pixels. The pixels are ranked according to these statistical values.

¹ http://www.sn.af.rl.af.mil/pages/SNH/ir_sensor_branch/example.html

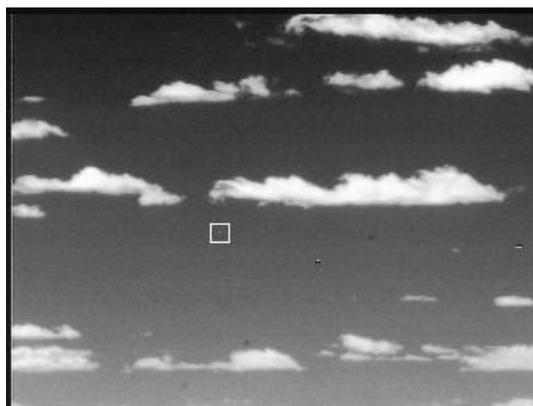


FIGURE 7. A sample frame from an infra-red video sequence. One-pixel target is shown inside the rectangle.

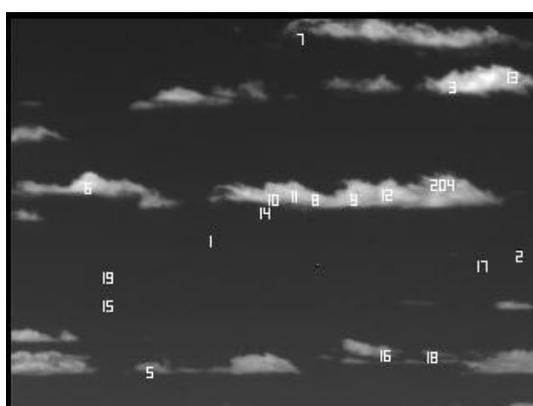


FIGURE 8. A sample output image, as a result of AWT followed fourth order test. The target is marked with one. The input images are frames 30 and 35 of sequence 1, detection results of which is shown in the third row of Table 1.

The next two methods do not include adaptive wavelet transform. The fourth order or FLO statistics are calculated around the local maxima pixels directly on the original image instead of the residual image.

The last method we employ does not include any filtering operation except the differencing. After the local maxima are extracted in the image, maximum and minimum valued pixels are found in 3 by 3 blocks. We rank the local maxima according to the value M_k defined as follows.

$$M_k = \frac{\max(e[k_x + i, k_y + j]) - \min(e[k_x + i, k_y + j])}{\min(e[k_x + i, k_y + j])} \quad i, j = -1, 0, 1 \quad k = 1..K \quad (8)$$

where K denotes the number of the local maxima detected in the image, and $e[i, j]$ is the pixel value at location (i, j) .

It is observed in Table 1 that adaptive wavelet based methods perform better than their counterparts without adaptive filtering. This is due to the fact that the significance of the edges of the target is amplified with the adaptive wavelet transform. While the adaptive wavelet transform and Higher Order Statistics (HOS) [14] based method performs best among other methods, the method based on Equation 8 is second to the best, since there is a significant difference between the maximum and minimum values in the image region containing a target.

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 Section 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 [2].

TABLE 1. Ranks assigned to point (one pixel) targets by various detection methods.

Sequence	Frame	Detection Methods						
		AWT HOS	&	HOS	AWT FLOS	&	FLOS	Loc.Max.
1	15	4		85		69	100	20
1	20	7		66		12	73	6
1	30	1		62		7	91	1
1	35	1		34		1	67	1
1	45	1		66		1	76	7
1	50	2		51		50	69	3
2	20	7		25		20	88	16
2	25	16		50		23	85	3
2	35	2		83		12	101	9
2	50	14		42		14	52	3
2	55	23		28		101	52	11
Averages:		7.09		53.82		28.18	77.64	7.27

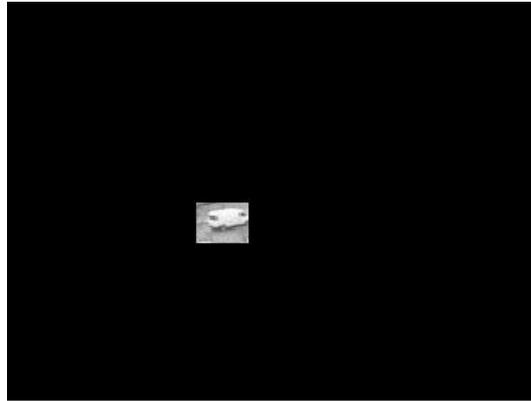


FIGURE 9. The detected moving object: Regions exceeding the threshold based on FLO statistic, geometric power.

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

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

where μ and σ_i are the mean and the standard deviation of the test statistic, respectively. The parameter λ is selected as 2.5 in our experiments. Normally it is selected as 3 as a rule of thumb which is based on the fact that if the test statistic has a Gaussian distribution than almost of the observations fall within the segment determined by the $3\sigma_i$. Anything exceeding the threshold $T_{h,i}$ is consider to be an outlier.

The image shown in Figure 7 shows the small regions exceeding the threshold based on the geometric power, the FLO statistic defined in Equation (7). The minivan shown in Figure 2 is clearly detected. The image shown in Figure 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. About 30 detection results are summarized in Table 1. In these detection experiments the number of false alarms for variance and geometric power based detection methods x per image and y per image, respectively. The use of geometric power significantly reduces the number of false alarms compared to the variance based detection method.



FIGURE 10. Regions exceeding the variance based threshold.

TABLE 2. Values of the test statistic $h(I_1, I_2, I_3, I_4)$ in regions with and without moving objects.

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

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 a detection trial 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]. Adaptive subband decomposition provides a good trade-off between 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 .

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 proposed method is also computationally efficient as the detection operation is carried out over quarter size images.

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. A. Ekin, A. M. Tekalp, R. Mehrotra, "Automatic extraction of low-level object motion descriptors," to be presented in *IEEE International Conference on Image Processing*, Thessaloniki, Spet. 2001.
4. 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. on Circ. and Systems for Video Tech.*, 9:(8), Dec. 1999.
5. Ö. N. Gerek, A. E. Çetin, "Linear/nonlinear adaptive polyphase subband decomposition structure for image compression" in *IEEE Trans. Image Processing*, October 2000.

TABLE 3. Detection performance of each method.

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

6. Omer N. Gerek, A. E. Cetin, 'Polyphase adaptive filter banks for fingerprint image compression,' Electronics Letters, pp. 1931-1932, vol. 34, Oct. 1998.
7. Omer Nezh Gerek, A. Enis Cetin, "Polyphase Adaptive Filter Banks for Fingerprint Image Compression" EURASIP Int'l. Conf. EUSIPCO'98, Sept. 1998.
8. R.Oktem, K. Egiazarian, A. E. Cetin, "Subband Decomposition Based Image Compression Algorithms With Nonlinear Adaptive Filter Banks", Proc. of IEEE- EURASIP NSIP 99, Antalya, Turkey, vol. 2, pp. 766-769, June 1999.
9. 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.
10. J. Gonzales, G. R. Arce, "Zero-Order Statistics: A signal processing framework for very impulsive processes," in IEEE Signal Proc. Workshop on Higher-Order Statistics, Banff, Canada, July 1997.
11. W. Sweldens. "The Lifting Scheme: A New Philosophy in Biorthogonal Wavelet Constructions", in Proc. of Society of Photo-Optical Instrumentation Engineers (SPIE), vol. 2569, pp. 68-79, Sept. 1995.
12. 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.
13. Rabi Zaibi, Yasemin Yardimci, A. Enis Cetin, "Small Moving Object Detection In Video Sequences," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2000), Istanbul, Turkey, June 2000.
14. Rabi Zaibi, Yasemin Yardimci, A. Enis Cetin, "Small moving object detection using adaptive subband decomposition in video sequences," in Signal and Data Processing of Small Targets 2000, Proc. SPIE Vol. 4048, p. 134-141, July 2000.