MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan

13-2

Two-Axial Subtraction of Estimated Background Integral Projections: Fast Method in Automatic Target Detection and Tracking

Ali Karandish
YMA College, Tehran, Iran
alikarandish@Gmail.com

Ali Mahmoodi
YMA College, Tehran, Iran
mahmoodia@Yahoo.com

Farid Behazin
YMA College, Tehran, Iran
behazin@Ieee.org

Abstract

In this paper we propose an Automatic Target Detection and Tracking (ATDT) method, which is appropriated for real-time implementation. In the proposed method, mapping of the image information from two-dimensional (2D) image domain to one-dimensional (1D) vertical and horizontal integral projections domain is used to decrease the under-processed data. An adaptive thresholding technique in target detection and also a flexible self-tuned algorithm in estimating the background representations are used in order to achieve in an accurate method. Based on aforementioned notation, a heuristic method is presented in this paper, which has adequate advantageous for real-time detection and tracking of a moving target(s) in an approximately stationary background with correlated and homogenous texture. We call this method as Two-Axial Subtraction of Estimated Background Integral Projections (TASEBIP).

1 Introduction

Automatic Target Detection and Tracking (ATDT) in image sequences is a complicated image processing application, which is prospected to be implemented in real-time [7,11]. The most principle reason of difficulty in implementation of such an ATDT application is the large amount of under-processed data, which is involved in an image sequence. Therefore, it is clear that decreasing the amount of under-processed data is a proper technique to achieve in such a fast ATDT method [3]. The amount of under-processed data is decreased too much, by mapping the image information from 2D image domain into 1D vertical and horizontal integral projections domain [5], in the proposed method. Mapping has been done by summation of image pixels to form integral projection representation of an image frame, which is intrinsically reducing the uncorrelated noise power and increasing *Signal to Noise Ratio* (SNR) in under-processed data set and then, increasing the potential of proposed method in detecting the target(s) within the noisy image sequences [11]. In addition, estimating the background integral projections via a suitable recursive routine is an adaptive method to increase the probability of the target detection and decrease the preprocessing and post processing steps in omitting the noise and clutter effects [2,8]. Verifying the results of the proposed method implementation and other known advantageous algorithms [6,8,9] indicates that TASEBIP is fast and accurate enough to be used as a real-time algorithm, especially in detection and tracking of the flying target(s) in *Forward Looking Infrared* (FLIR) image sequences [10].

2 Integral Projections

Integral Projections (IPs) of the image frame have been used to reduce the amount of under-processed data. These image representations (Vertical and horizontal IP) can be formed easily through the equation pairs (1) and (2), and have been shown in Figure1.

$$\left.\begin{aligned}\mathbf{HIP}_k &= \{\mathbf{HIP}_k(x) \mid x = 1,2,\ldots,M\}\\ \mathbf{VIP}_k &= \{\mathbf{VIP}_k(y) \mid y = 1,2,\ldots,N\}\end{aligned}\right\} k = 1,2,\ldots,n \qquad (1)$$

$$\begin{aligned}\mathbf{HIP}_k(x) &= \sum_{i=1}^{N} I_k(x,i)\\ \mathbf{VIP}_k(y) &= \sum_{i=1}^{M} I_k(i,y)\end{aligned} \qquad (2)$$

M and N are the number of rows and columns of pixels in each image frame; k is the frame number, $I_k(i,j)$ is the gray level values of the pixel at row i and column j of frame k, and n is the number of frames in the sequence.

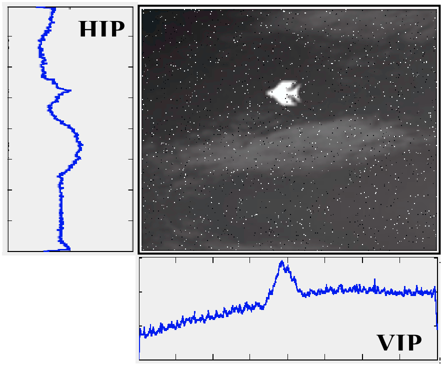


Figure 1. Vertical and Horizontal Integral Projections (VIP & HIP) of an image frame

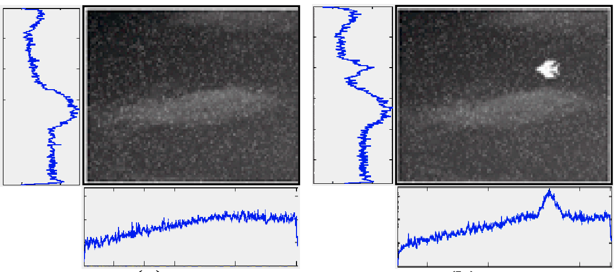
Because of the recognizable similarity between $\mathbf{HIP}$ and $\mathbf{VIP}$ from the processing and formulation point of view, we just present the equations formulated for $\mathbf{HIP}$ in the subsequent part of this paper. However, in all cases the equivalent equations for $\mathbf{VIP}$ can be achieved by substitution of $\mathbf{H}$, M and x by $\mathbf{V}$, N and y sequentially.

3 Target Detection

If it's considered that Integral Projections of Background for previous frame ($\mathbf{BHIP}_{k-1}, \mathbf{BVIP}_{k-1}$) are estimated properly, which is discussed in details later, the appeared target would be detected easily in each frame using simple one dimensional point to point subtraction through the equation (3) as follows [11].

$$\mathbf{HIP}_k \quad \left|\mathbf{HIP}_k^* - \mathbf{BHIP}_{k-1}\right| \qquad (3)$$

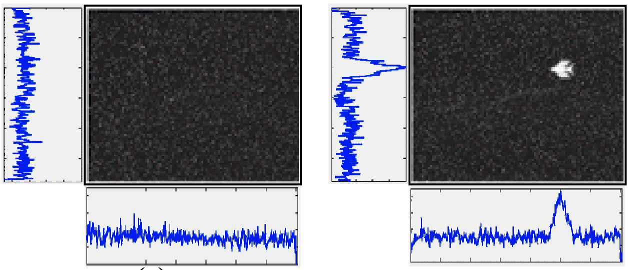
For example IPs of two different frames of an image sequence, are shown in Figure2.



(a) (b)

Figure 2. Two different frames of an image sequence with their VIPs and HIPs. (a) Image frame contains background only and (b) Image frame contains the both background and target

Figure3 illustrates what has meant here by subtraction of IPs to detect target. There would be just a noisy flatted IPs for frames which contain background only, but there would be some recognizable peaks in IPs of frames which contain the both target and background.



(a) (b)

Figure 3. Subtraction diagrams of HIP and VIP of frames in Figure 2 and estimated background VIP and HIP

The IPs of the separate target would be extracted by thresholding the subtracted IPs. Thresholding levels ($\Im_H$, $\Im_V$) can be assumed as fixed constants for all frames or can be calculated adaptively for each frames individually [1,4,10]. Here, an effective semi-empirical routine proposed to find the threshold levels adaptively. It's possible to normalized subtracted IPs of all frames and reorder the values between 1 and Naperian constant (e) using the equation (4).

$$\Delta \mathbf{HIP}_k^{(e)} = \exp\left(\frac{\Delta \mathbf{HIP}_k}{\max\{\Delta \mathbf{HIP}_k\}}\right) \tag{4}$$

If $\mu_H(k)$ and $\sigma_H^2(k)$ are the mean value and variance of $\Delta \mathbf{HIP}_k^{(e)}$ array, then the threshold values can be calculated adaptively through the equation (5) for each frame of any sequence [10].

$$\Im_H(k) = \mu_H(k) + \frac{a_1}{\sigma_H^2(k)} - a_2 \tag{5}$$

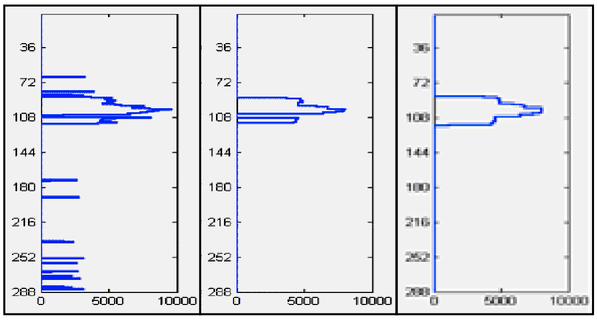
Where a_1 and a_2 are positive constants, which will be determined at ATDT system start up, based on environmental features of imaging scene [10]. Note to adaptive thresholding routine, the thresholded IPs of each frame is formed in according to the equation (6).

$$\Delta \mathbf{HIP}_k^{\Im_H}(i) = \begin{cases} \Delta \mathbf{HIP}_k^{(e)}(i); & \Delta \mathbf{HIP}_k^{(e)}(i) \geq \Im_H(k) \\ 0 \quad ; & otherwise \end{cases} \quad i = 1,..,M \tag{6}$$

One-dimensional Min-Max filters have been used to omit noise and clutter interference from thresholded IPs. Min-Max filters are defined as the equations (7) and (8), which have effects as similar as well known 2D Closing and Opening morphological filters. The effects of aforementioned filters are shown in Figure4 step by step.

$$\Delta \mathbf{HIP}_k^{\nabla} = \min_{i=2}^{M-1}\left\{\Delta \mathbf{HIP}_k^{\Im_H}(m) \mid m = i-1, i, i+1\right\} \tag{7}$$

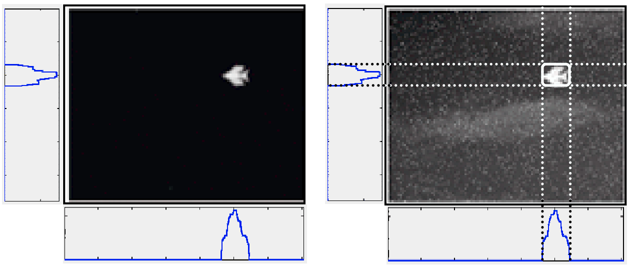
$$\Delta \mathbf{HIP}_k^{*} = \max_{i=3}^{M-2}\left\{\Delta \mathbf{HIP}_k^{\nabla}(m) \mid m = i-2, \dots, i+2\right\} \tag{8}$$



(a) (b) (c)

Figure 4. Omitting the noise and clutter effects in thresholded result of an instant subtracted IP, using 1D Min-Max filters. (a) Thresholded IP before filtering, (b) Result of using Min filter in a 3 point neighboring on thresholded IP, (c) Result of using Max filter in a 5 point neighboring on filtered IP, which has been shown in (b)

Using what has been explained herein, the IPs of image frame will be thresholded and filtered and finally the target will be detected as peak locations in what remains from frame IPs. The result of such a task is shown in Figure5-a. It's clear that the image frame will be segmented and the target position will be determined by expanding the widths of the target IPs over image frame and to form the target bounding box properly (Figure5-b).



(a) (b)

Figure 5. (a) Image frame Integral projections of Figure3-b after thresholding and Min-Max Filtering and (b) Image segmentation and the target detection using the subtracted IPs after thresholding and also filtering

It's also possible to determine the centroid of the target with an appropriate accuracy using the equation (9).

$$c_x(k) = \frac{[1 \quad 2 \quad 3 \quad \dots \quad M] \times \left(\Delta \mathbf{HIP}_k^{*}\right)^T}{\sum \Delta \mathbf{HIP}_k^{*}} \tag{9}$$

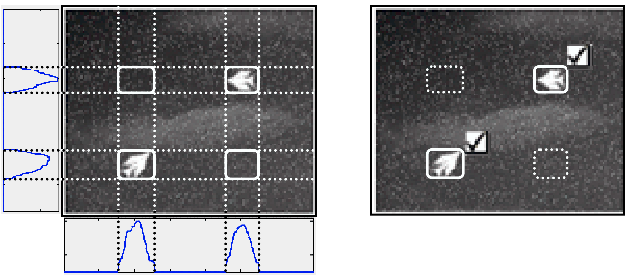
Here, T evokes the involved target IP arrays transposed.

1.1 Multi-target detection

With some suitable modifications TASEBIP would be well adequate in the case of Multi-Target Detection. In such a condition, it's comprehensible that by following the basics of TASEBIP, there may be more than a peak in what remains from the IPs, based on the topology of the targets in each frame. For an instance, Figure6-a demonstrates an image frame of a sequence contains two targets with its IPs after thresholding and filtering. In this case, both of $\Delta\mathbf{HIP}_k^*$ and $\Delta\mathbf{VIP}_k^*$ contain two peaks, and therefore, expanding the widths of these peaks over the image frame results in segmenting four individual regions that each of them may or may not contain any target. In order to omit the invalid regions, an effective routine is proposed hear, which is called as *Segment Validation Routine* (SVR). In this routine, the absolute difference between gray level Mean-Values of pixels in each segments (μ_s) and pixels in an equal-sized part of background (μ_b) verified with a threshold value ($\Im_s$), which will be tuned based on minimum difference between background and an observable target intensity Mean-Values in the application case [10]. If the difference reach the threshold, the involved segment is validated, else it is omitted formally. If $L_s(r)$ is assumed as the label of segment r, then SVR can be defined in according to the equation (10).

$$L_s(r) = \begin{cases} Target & ; \quad |\mu_s - \mu_b| \geq \Im_s \\ Non-Target & ; \quad otherwise \end{cases} \tag{10}$$

Figure6-b illustrates the final effect of SVR in labeling the valid segments in mentioned example.



(a) (b)

Figure 6. (a) Result of using TASEBIP method in the case of detecting multi target in an image frame, (b) Effect of using SVR in labeling the valid segments

1.2 Multi-target tracking

Status prediction procedure of the targets is also an essential stage in *Multi Target Tracking* (MTT) applications, in order to discriminate the detected targets during tracking process. In this case, the method of prediction is based on the level of similarity between the targets and their motion characteristics. For an instance, it's perceived that third order polynomial extrapolation of position is effective enough to discriminate the targets accurately, from tracking process of twenty tested infrared image sequences of at least two and at most five simultaneous flying objects [10].

4 Estimating of Background IPs

If it's allowed to have frame(s) which doesn't contain any target at the beginning of the under-processed image sequences, the estimated Background Integral Projections ($\mathbf{BHIP}$, $\mathbf{BVIP}$) would be easily set to the first frame smoothed IPs through the equation (11).

$$\mathbf{BHIP}_1 = \mathbf{HIP}_1^* \tag{11}$$

According to the explained procedure of the target(s) segmentation, the background IPs in the coming frames of the sequence can be estimated and adaptively updated following the equations (12), (13), (14) and (15) consecutively [10].

$$\mathbf{HM}_k = \left\{ \mathbf{HM}_k(i) \middle| \mathbf{HM}_k(i) = \begin{cases} 1; \Delta\mathbf{HIP}_k^*(i) = 0 \\ 0; \Delta\mathbf{HIP}_k^*(i) \neq 0 \end{cases}, i = 1,.., M \right\} \tag{12}$$

By using $\mathbf{HM}_k$ as the matched target(s) IP rejecting mask, we have:

$$\mathbf{BHIP}_k^{\nabla} = \mathbf{HM}_k \bullet \mathbf{HIP}_k^* + \overline{\mathbf{HM}_k} \bullet \mathbf{BHIP}_{k-1} \tag{13}$$

Note that the $\bullet$ sign shows the internal point to point multiplication and also:

$$\overline{\mathbf{HM}_k} = \mathbf{Not}(\mathbf{HM}_k) \tag{14}$$

Finally, the estimation of background IPs can be calculated adaptively for every frame of the image sequence according to the equation (15) during ATDT process.

$$\mathbf{BHIP}_k = \alpha \times \mathbf{BHIP}_{k-1} + (1-\alpha) \times \mathbf{BHIP}_k^{\nabla} \quad , k = 2,3,...,n \tag{15}$$

In equation (15), α is the refreshing factor ($0 < \alpha < 1$), and is determined based on background variation rate, the camera movement characteristics and the target(s) speed, relatively [10].

5 Performance Evaluation

Performance of proposed method has been evaluated by implementing an ATDT system based on TASEBIP method. For estimating the performance of proposed ATDT method, error value between the response of implemented system in determining the position target and the accurate position value is calculated for each frames of a simulated image sequence. Then the standard deviation and Mean-Value of errors (RMSE) determined for the mentioned image sequence. This test has been repeated for the same image sequence with different amounts of uncorrelated noise.

Due the fact that SNR value affects fundamentally on the system performance, it's conventional to evaluate the error versus Mean-SNR to achieve in performance criteria. In evaluation process, a modified type of Mean-SNR value is used which is formed as follows. If $I_{Tgt}(k)$ assumes as the image of the target in frame k of a sequence, the equivalent energy of the target in the image space ($E_{Tgt}(k)$) is calculated as the observed signal according to the equation (16).

$$E_{Tgt}(k) = \sum_{\rho_{xy} \in I_{Tgt}(k)} \rho_{xy} \tag{16}$$

In equation (16) ρ_{xy} is the gray level value of a pixel at position (x,y) in an image frame. Also the noise equivalent energy ($E_{noise}(k)$) can be determined as follows:

$$E_{noise}(k) = \mathrm{var}\left(I_{noisy}(k) - I_{noiseless}(k)\right) \tag{17}$$

Modified SNR value of each frame ($SNR_M(k)$) is declared by the equation (18).

$$SNR_M(k) = 10 \times \log \frac{\sum_{\rho_{xy} \in I_{Tgt}(k)} \rho_{xy}}{\mathrm{var}(I_{noisy}(k) - I_{noiseless}(k))} \quad (d\beta) \tag{18}$$

Finally the Mean-SNR value of an image sequence is determined as follows:

$$\tilde{SNR} = \frac{1}{m} \sum SNR_M(k) \quad \forall k \mid I_{Tgt}(k) \subset I_{noisless}(k) \tag{19}$$

Note that m is the number of the frames which the target appears in, over an image sequence. The performance criteria of TASEBIP have been shown in Figure8. These criteria have been formed based on the result of TASEBIP in detection and tracking of a flying target in some infrared image sequences with different amount of noise, ambient structure and characteristics.

Using a PC system with a Intel® Pentium IV CPU (2020MHz), 523744 kilobytes of total physical memory and Microsoft Windows 2000 Professional™ as operational system, it is possible to implement TASEBIP even in MATLAB® (Version 6.5.0.180913a - Release 13) as a real time ATDT system to detect and track target(s) in the sequences with frame size of 288x352 pixels. Results of implementation indicate that processing time of each frame is approximately 32 milliseconds in average [10].

The ATDT system based on TASEBIP method has been tested on 200 different infrared image sequences of flying object(s) with image size of at least 5x5 pixels, containing both simulated and real ones, and has completely succeeded in tracking the target(s) in the sequences with $\tilde{SNR}$ value more than $+30\,d\beta$ [10].

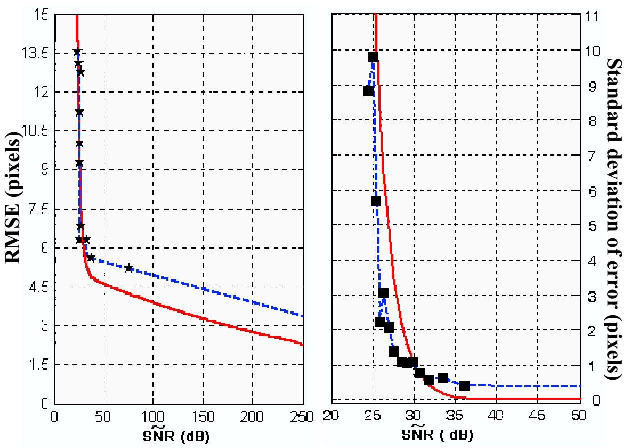


Figure 8. Performance criteria of TASEBIP

RMSE and standard deviation of error versus $\tilde{SNR}$

6 Conclusion

In this paper, we proposed an appropriate method for ATDT application, which has been called as TASEBIP. The proposed algorithm formulation and its procedure have been described in details. Integral Projections (IPs) of the image frame is playing a basic role in this algorithm. Mapping the two dimensional image domain into the one dimensional IPs domain, with the goal of decreasing the amount of under-processed data set has been used to form a fast algorithm. An adaptive thresholding technique in target detection and also a flexible self-tuned algorithm in estimating the background IPs, have been resulted in an accurate method. Efficiency of the proposed algorithm in single target and multi target detection and tracking tasks has been fundamentally investigated. Performance criteria of the proposed method have been presented in this paper. The experimental results have indicated that this method is fast and accurate enough to be used as a real time ATDT algorithm.

References

[1] J. Boriotti, "Automatic Thresholding for Detection of Landing Aircraft in FLIR Imagery," *Proc. SPIE*, vol. 1957, pp. 156-159, April 1993.

[2] M.J. Carlotto, "Nonlinear Background Estimation and Change Detection for Wide-Area Search," *Optical Engineering,* vol. 39, no. 5, pp. 1223-1229, May 2000.

[3] K.L. Chung, and L. Chang, "A New Predictive Search Area Approach for Fast Block Motion Estimation," *IEEE Trans. On Image Processing,* vol. 12, no. 6, pp. 648-652, June 2003.

[4] W. Lie, "Automatic Target Segmentation by Locally Adaptive Image Thresholding," *IEEE Trans. on Image Processing,* vol. 4, no. 7, pp. 1036-1041, July 1995.

[5] J.S. Kim, and R.H. Pak, "A Fast Feature Based Block Matching Algorithm Using Integral Projections," *IEEE Journal; Selected Area In Communications,* vol. SAC-10, pp. 968-971, June 1992.

[6] H. Lange, "Real-Time Moving Target Detection Based on a Nonlinear Filter Using Short Time And Medium Time Image Differences," *Part of SPIE Conf. on Signal Processing, Sensor Fusion, and Target Recognition VIII,* Orlando, FL, vol. 3720, pp. 202-213, April 1999.

[7] A. Howard, C. Padgett, and K. Brown, "Real Time Intelligent Target Detection and Analysis with Machine Vision," *Proc. of the 3rd International Symposium on Intelligent Automation & Control, World Automation Congress,* June 2000.

[8] E. Rivlin, M. Rudzsky, R. Goldenberg, U. Bogomolov, and S. Lepchev, "A Real-Time System for Classification of Moving Objects," *in Proc. Of IEEE ICPR'02,* Quebec City, Canada, vol. 3, pp. 688-691, August 2002.

[9] H. Veeraraghavan, O. Masoud and N. Papanikolopoulos, "Real-Time Tracking for Managing Suburban Intersection," in *Proc. Of IEEE DSP'2002,* vol.2, pp.1023-1026, Greece, 2002.

[10] A. Karandish, ***Real Time UFO Tracking & Classification in Thermal Image Sequences,*** Electro Optical Eng. M.S. Dissertation, MUT, Iran, September 2003.

[11] G. Wang, ***A Parallel System for Single-Pixel Target Detection and Tracking based on Trajectory Continuity Theory,*** Electrical Eng. Ph.D. Dissertation, University of Virginia, August 1991.