A COMPARISON OF VISUAL TARGET TRACKING METHODS IN NOISY ENVIRONMENTS

M. Sezgin, S. Birecik, D. Demir, İ.Ö.Bucak, S. Çetin, F. Kurugöllü

TUBITAK Marmara Research Center , Department of Optoelectronics, PK 21, 41470, Gebze, Kocaeli-TURKEY
phone: +90-262-6412300; fax: +90-262-6412309; e-mail: sezgin@yunus.mam.tubitak.gov.tr

Abstract - This work represents a comparison of several methods for target tracking in noisy environments using motion-energy and template location based approaches. In order to remove the registration noise, a morphological filter has been adopted. Since the scale and rotation of the target change, a template update method has been used. The performance of different methods to compare their respective speed and accuracy have been evaluated and the results obtained are presented.

I. INTRODUCTION

Motion detection and tracking are interesting and useful area in computer vision. They can be used in military and industrial applications. There have been many works concerning the detection and tracking problem. Here, we compare several methods for target tracking in noisy environments using motion-energy and template location based approaches.

Motion-energy based target tracking systems rely entirely on motion detection to detect the moving object. They have the advantage of being able to track any moving object regardless of size or shape [1]. In addition to computational simplicity, motion-energy method is suitable for pipeline architectures, which allow it to be readily implemented on most high-speed vision hardware. The obvious disadvantage is that it is not robust against noise.

In template location based methods, a template image is searched in a given scene. They have the advantage of giving the better results in noisy environments, but it takes longer time [2]. Four methods for image location are outlined in this work. The performance and robustness against noise of these algorithms have been tested using real images with additive noise.

II. MOTION-ENERGY METHOD

In this method, an image is segmented into regions of motion and inactivity. The block diagram of the method is described in Fig. 1. The simplest implementation of motion-energy detection is image subtraction. Consecutive images I(t-1) and I(t) are subtracted from each other, pixel by pixel. Fig. 2 (a) and (b) shows two frames of an image sequence taken with a CCD camera. The absolute value of the difference image is taken and thresholded by moment-preserving [3] method to segment the image into static and dynamic regions. The illumination and sensor noise is eliminated by thresholding. Fig. 2 (c) shows the result of image subtraction and thresholding.

Since registration noise will definitely occur; it usually consists of thin lines that may seriously influence the calculation of the target's location. In this work , morphological filtering is used to remove it. Fig. 2 (e) shows the result of morphological filter. In order to estimate the position of the moving target at time t, a binary edge image of the current frame is obtained. Here, Sobel edge detector is used. Fig. 2 (d) shows the result of Sobel edge detector. Then this image is ANDed by subtracted and filtered image. At the last stage, centroid of the moving target image is calculated. Fig. 2 (f) shows the result of last step.

Morphological filtering

Morphological filters [4], [5] have been applied to the images for removing registration noise. Registration noise can not be removed via the thresholding operation exactly. Here, we used the opening operation. Opening of a binary image A by a structuring element B, denoted by O(A,B), is defined as an erosion followed by a dilation.

$$O(A,B) = D\,[E(A,B),B] = [A\Theta(-B)] \oplus B \qquad (1)$$

where Θ and $\oplus$ denote Minkowski subtraction and addition, E and D denote erosion and dilation operations respectively.

By applying erosion to the difference image, narrow noisy regions can be eliminated. Since the width of registration noise is usually one or two

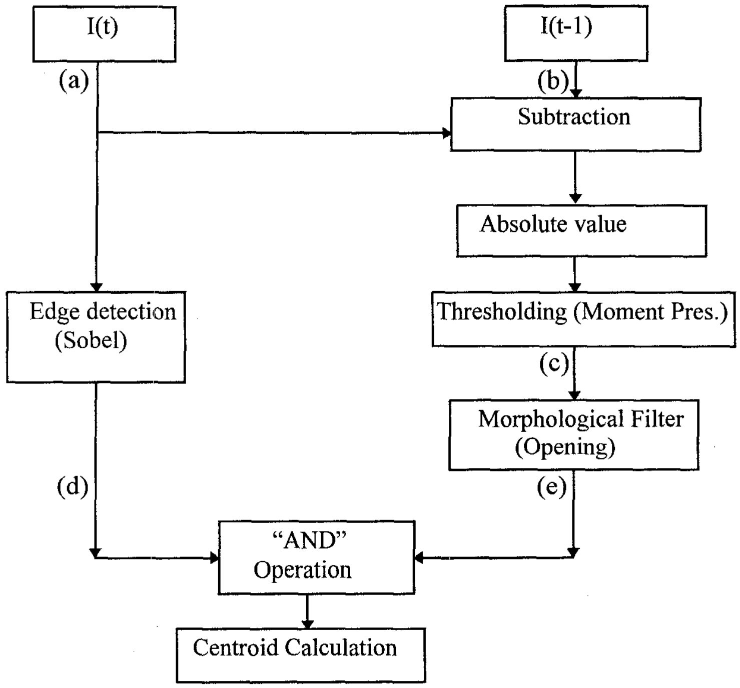

Fig. 1 Block diagram of motion-energy method.

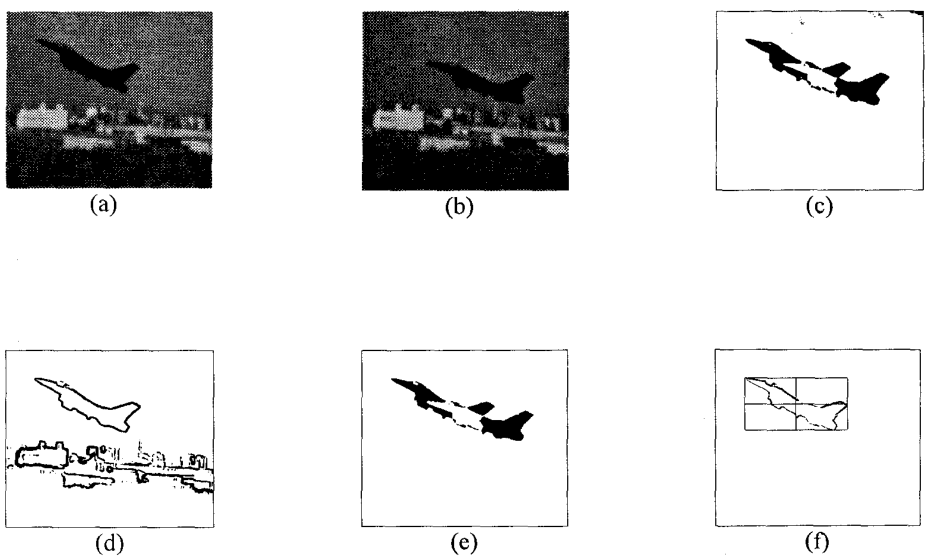

Fig. 2 Results of algorithmic steps in motion-energy method,
a) The image observed at time t: I(t), b) The image observed at time t-1: I(t-1), c) Subtracted and thresholded image, d) Edges of I(t) by using Sobel operator,e) Result of the morphological filtering, f) Centroid of the moving target after "AND " operation.

pixels, 2x2 or 3x3 square structuring elements shown in Fig. 3 is used. The extracted moving target image will be thinned, not completely eliminated. However, it can be recovered by another dilation operation on the eroded moving target image with the same structuring element. If the erosion mask is wider than a given region that region will be eliminated completely and not appear after dilation.

Fig. 3 The structuring elements used in noise filter

III. TEMPLATE LOCATION METHOD

Motion-energy method does not give better results in noisy environments. In that case, template location techniques are more effective. In this part, the performance of different template location methods[2] have been evaluated and results obtained are presented. Here, we used the following notation: I(k,l): KxL, is the image in which we want to find a certain image, and T(i,j): NxM, is the template image. Fig. 4 shows the template image of size 25x61 pixels.

Fig. 4. Template image: T

Template matching 1: Sum of the absolute values of difference between the pixels of the picture and the window

The objective of the algorithm is to minimize the following distance function [2]:

$$\min_{(k,l)}\{d1(k,l)=\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\left| I(i+k,j+l)-T(i,j)\right|\} \quad (2)$$

If the size of the images are relatively large, the execution time of the algorithm is greater than that of methods based in the use of Fourier space. However, it has the advantage of using only integer operations. Therefore, this is one of the most widely used real-space based methods.

Template matching 2: Sum of the squared values of difference between the pixels of the picture and the window

This method defines the following distance function [2]:

$$\min_{(k,l)}\{d2(k,l)=\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}(I(i+k,j+l)-T(i,j))^2\} \quad (3)$$

The distance between two images is then defined as the sum of the squares of the differences between their pixel amplitudes. As in the previous method, the position with the minimum cost function is selected as the location of the desired pattern.

Template matching 3: Cross-correlation method

This method has been the most widely used for object detection in noisy images. The most probable location for a template in an image is the one that maximizes the following function:

$$\max_{(u,v)}\{d3(u,v)=\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}I(i+u,j+v)T(i,j)\} \quad (4)$$

This method can be implemented very fast way by using FFT algorithms. In order to get two FFTs with the same amount of points, it is necessary to introduce the template image of size M xN in an array of size KxL, filling the rest of the array with zeros. Maximizing the cross-correlation function is functionally equivalent to minimizing the sum of the squares of the differences of the pixels between two images in the case of pictures whose energy is approximately constant over windows of size MxN.

Template matching 4: Correlation coefficient method

When the energy of the image is not constant over windows of size M, it is necessary to normalize the cross-correlation in order to get comparable values, independent of the picture energy. For that reason, the correlation coefficient is used. It is defined as

$$\max_{(u,v)}\{d4(u,v)=\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}I(i+u,j+v)T(i,j)}{\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}I^2(i+u,j+v)\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}T^2(i,j)}}\} \quad (5)$$

The numerator is equivalent to the cross-correlation function. For this reason, it can be calculated using the FFT algorithm.

Template update

If there is a rotation or scaling change in the template , constant template hypothesis does not give good result. To prevent this deficiency, dynamic region which is obtained from last step of the motion energy method is used to update template. For this purpose inner region of rectangular window in Fig.2-f was used to update template. If mean and variance of the rectangular window pixels gives a difference more than $\pm 10\ \%$ template is updated.

IV. EXPERIMENTAL RESULTS

In this section, the different methods for target tracking have been tested on real images. The hardware platform consists of a Pulnix TM765 CCD camera, Scorpion real time frame grabber (640x480 and 8-bit resolution) and 486DX2-50 PC. Algorithms have been run in Borland C software. First, noise performance has been given. The noise type used for this purpose is additive Gaussian noise. Fig. 5 and Fig. 6 show the results of the application of different algorithms on noisy images of size 160x160 with SNR=3 dB and SNR=1 dB. Since we have discarded the cross-correlation method as useless in noisy environments, the result of the application of this method is not shown. As shown, motion-energy method did not work well in SNR=1 dB.

Table-1 shows the execution times of the algorithms. After a close examination of the Table 1, one observes that motion-energy method is faster than template location methods.

Fig. 7 shows the location error with respect to SNR. The location error has shown oscillation when the SNR was increasing.

V. CONCLUSIONS

Several methods have been described and tested for target detection. It has been observed that motion-energy method is the fastest of all the methods described. But it is not robust against noise. In order to remove sensor and illumination noise, moment-preserving method has given better results. Cross-correlation method has shown the worst performance in detecting the position of targets. Correlation coefficient method has given the best results in noisy environments.

In the future, a novel video tracking system with adaptive predictors are planned to be developed.

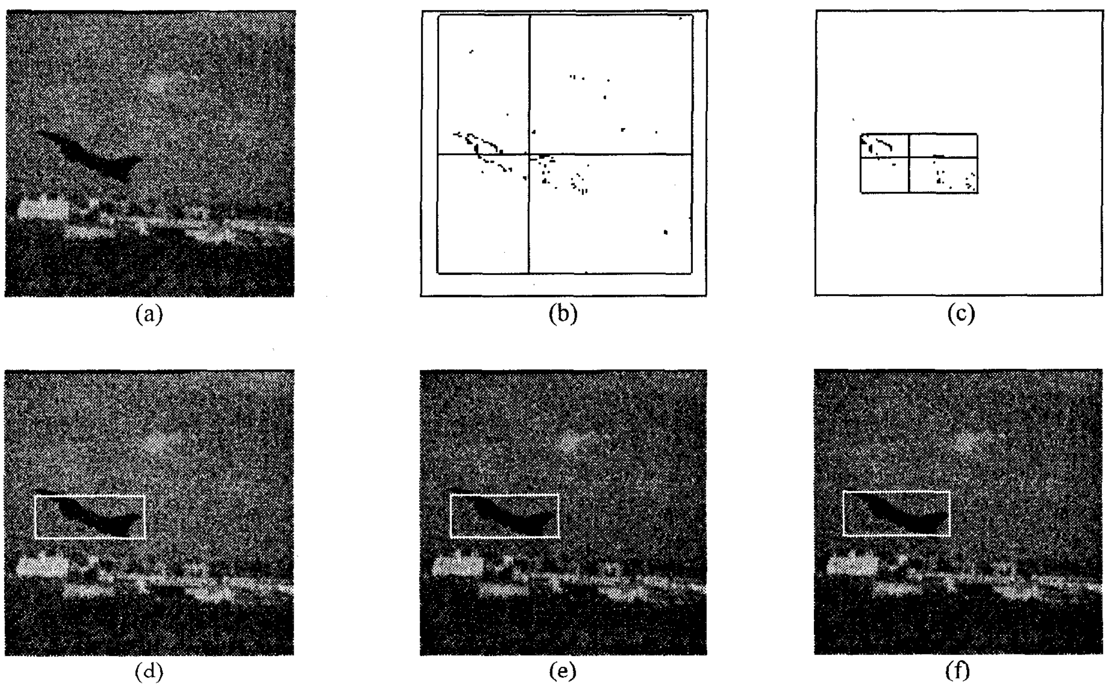

Fig. 5 Results of the different algorithms for SNR=3dB,
a) noisy image, b) motion-energy for 2x2 structuring element, c) motion-energy for 3x3 structuring element, d) template matching-1, e) template matching-2, f) template matching-4.

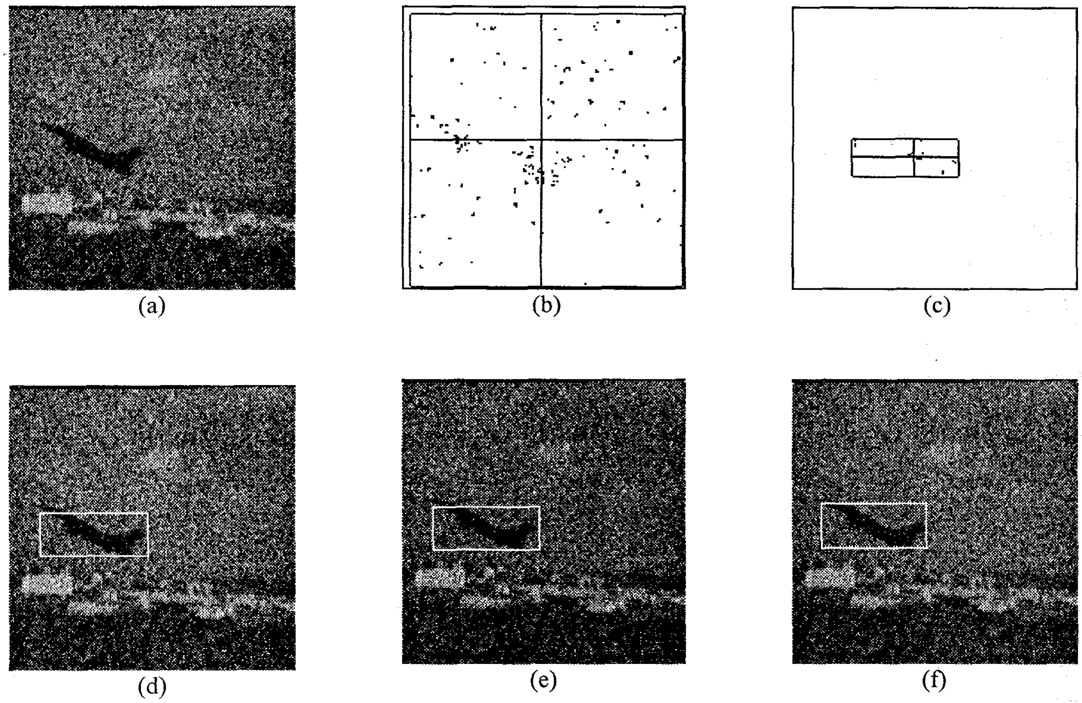

Fig. 6. Results of the different algorithms for SNR=1dB,
a) noisy image, b) motion-energy for 2x2 structuring element, c) motion-energy for 3x3 structuring element, d) template matching-1, e) template matching-2, f) template matching-4.

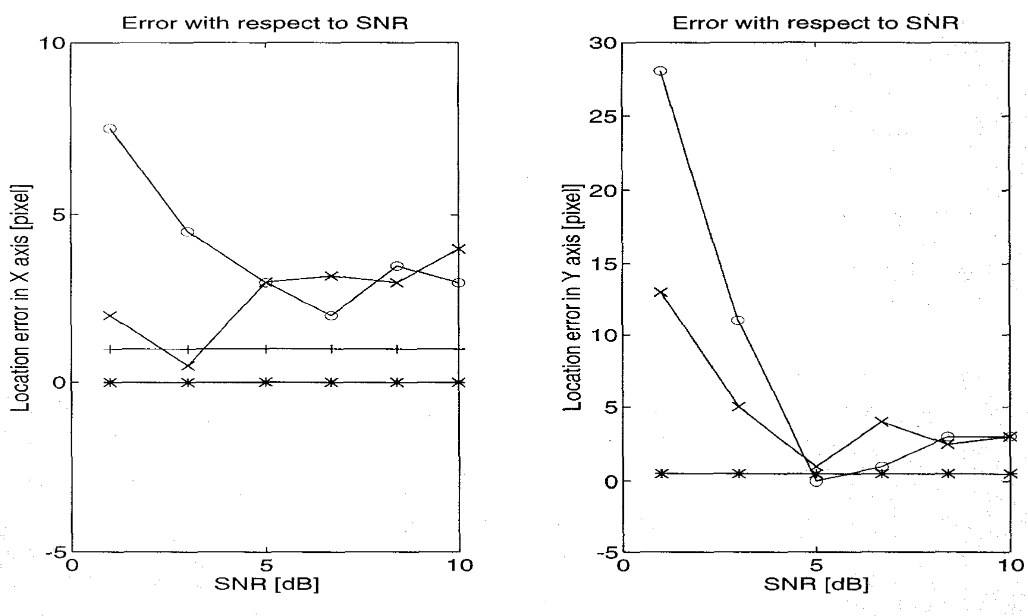

Fig. 7 . Position error with respect to noise in X and Y axis ,
o:Error of the motion energy method with 3x3 structural element ,**x**: Error of the motion energy method with 5x5 structural element , +:Error of the template matching -1, -:Error of the Template matching -2, *:Error of the template matching -4 .

Table-1 Execution time of the methods

Method	Motion Energy	Template Matching-1	Template Matching-2	Template Matching-4
Time (sec)	6	104	142	35

REFERENCES

[1] D. Murray, A. Basu , "Motion tracking with an active camera", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol:16, no:5,pp:449-459, 1994 .

[2] J.P. Secilla, N. Garcia ,"Template location in noisy picture", *Signal Processing*, vol:14,pp:347-361, 1988.

[3] P.K. Sahoo et. al.,"A Survey of Thresholding Techniques", *CVGIP*, vol:41, pp:233-260, 1988.

[4] L.H. Chen , S. Chang, "A video tracking system with adaptive predictors" *Pattern Recognition*, vol:25, pp:1171-1180, 1992.

[5] C.R. Giardina, E. R. Dougherty, *Morphological Methods in Image and Signal Processing*, Prentice Hall, 1988