Modified Hausdorff Distance Transform Technique for Video Tracking

Sarbajit Pal and P.K.Biswas[+]

Variable Energy Cyclotron Centre, Department of Atomic Energy
1/AF Bidhannagar, Calcutta 700064
sarbajit@veccal.ernet.in

[+]Department of Electronics and Electrical Communication Engineering
IIT, Kharagpur - 721302, India
pkb@ece.iitkgp.ernet.in

Abstract

The recent advances in computer and DSP technology have brought new expectation as well as new challenges in real time visual tracking of moving targets from video sequences. Due to the processing time constraint and continuous change in target size, segmentation and feature extraction for matching becomes very unreliable. In this paper we have presented an automated correlation based tracking approach using edge strength and Hausdroff Distance Transform (HDT) technique for tracking of moving targets. The system works in three states, Locking, Tracking and Recovery State. Difference image processing technique or manual selection is used in locking state to identify the moving target. Tracking state tracks the identified target over subsequent video frames. The system has been successfully tested with various real video sequences. The tracking performance evaluation and comparative results using Sobel's edge and HDT template are presented considering the image sequence where object is under high background noise condition.

Key words: Video Sequence Processing, Target Recognition, HDT, Template Matching.

1 Introduction

The recent advances in computer and DSP technology have brought new expectation as well as new challenges. A real-time tracker can be implemented using a computer with a real-time video grabber add-on card. This system can interact with dynamic world by analyzing digitized images to determine information about scene or object in the real world. There are several computer vision and image processing techniques to aid in this task, attaining success in limited domain.

The autonomous detection and tracking of moving objects is an important requirement in many fields spanning military, industrial and biomedical applications. Examples are detection of missile or airborne object from a surveillance post and detection of airborne targets from air.

The tracking of the objects can be categorized in two classes. One is tracking with stationary camera and second is tracking with moving camera. In case of moving cameras the movement of the camera can be controlled manually or automatically. However tracking with manually controlled camera can be used in applications where the object motion between successive video frames is only a small fraction of the camera field-of view (FOV). In such situations, off-line technique of determining exact location of the object within the camera FOV can also be used if the object remains within the FOV of fixed camera as in several industrials and biomedical applications.

However when tracking objects, such as missiles or aircraft, the object is likely to move out of the camera FOV and it is not feasible to track the object manually as the rapid motion between successive video frames becomes a limiting factor in such cases. Therefore it is required to implement automatic camera orientation control for motion tracking using image-processing technique in real time. Availability of high-speed image processing hardware and efficient algorithms have made possible to attempt designing a real time video tracking system. To make the system real time the image captured by the camera must be processed before the next frame digitized.

2 Previous works

Most of the previous work in the detection or tracking of object of interest in the scene can be classified into three categories, i.e. Correlation based tracking, Optical flow based tracking and Feature based tracking.

Although there have been several object tracking system proposed in the past, very few of these have actually addressed the concerned problems of real world. Such as occlusion, shadow, cloud cover, haze, seasonal variation, clutter, moving background and various other form of image degradation where recognition has to be accomplished under varying contrast and presence of shadow.

MPEG-4 Video coding techniques [1][3] that detect motion or use motion to extract feature of the target, rely on the constraint that the movement of the target is small between successive images allowing the object to be tracked using its spatio-temporal continuity. Constraining the movement to ensure overlap between two successive images puts restrictions on speed of the object.

Gilbert et. al. [2] proposed statistical classification based approach for moving object detection.

The optical flow based tracking [4] schemes are capable of giving accurate location of moving object but these schemes are computationally expensive and suitable for stationary background conditions.

In decision theory algorithm [5] the classification of pixels into stationary, moving and uncovered background is necessary. The stationary background requirement is rarely feasible in real world situations.

The dynamic programming algorithm [6] identifies targets by performing the equivalent of an exhaustive search of all possible target state sequences representing physically realizable target path, thus makes the system computationally expensive and unsuitable for real time applications.

The correlation based tracking algorithms [7] computes the noise statistics, simulate the target pattern and compares with corresponding threshold. This method involves extensive computation of covariance matrix and its inverse or Fourier phase difference.

Motion estimation based video tracking techniques have been reported by Reid et. al.[8]. Here the detection of motion can be performed by simple operation such as image differencing followed by thresholding. This method is attractive from computational viewpoint. But its application is limited to scenarios where there is absolutely no background motion.

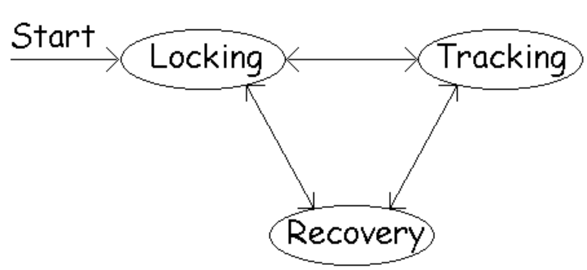


Fig. 1, System States

3 Proposed Scheme

The proposed scheme of object tracking system specifically works for outdoor sequences. The system works with thermal as well as optical image sequences and able to work without using model object. It is assumed that targets follow a particular trajectory. There are no abrupt changes in the target path. Speed of the target is constant or increasing gradually. There are no abrupt changes in the speed and target intensity is uniform.

The system is able to detect and track target against moving background. The movement in background is due to actual movement in the background or continuous movement of camera or due to noise and other disturbances. The system is designed for very small and variable taget size. During the tracking phase the target may be temporarily lost as it goes behind the clouds or out of FOV of camera. In such cases the system will recover the target by manual target detection.

The system can be used in three modes. In Automatic Detection the system will find out the target automatically and start tracking the same. Semiautomatic Detection is designed in keeping the view of multiple targets where user can select the target to be tracked.

4 Implementation Detail

The proposed video tracking system is a combination of segmentation and correlation based approach. The main novelty of our system lies in its use of simple template for object detection and tracking. The system is modeled as a three-state machine as shown in Fig.1.

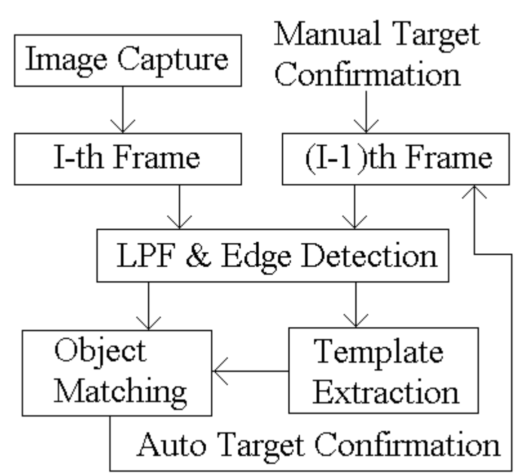


Fig. 2, Schematic Diagram of the tracking system after initial target selection

This has three states, Locking, Tracking and Recovery. Functions of each state are as follows. During the locking state dense motion estimation, segmentation and scene cut detection are carried out. To avoid false motion indication each individual frame is low pass filtered to remove noise before difference image frame is carried out. The underlying principle is to take difference of two consecutive video frames. The two frames $F_k(x,y)$ and $F_{k+1}(x,y)$ captured by camera at the time t_k and time t_{k+1}, are modeled as follows.

$$F_k(x,y) = S_k(x,y) + M_k(x,y) + N_k(x,y)$$
$$F_{k+1}(x,y) = S_{k+1}(x,y) + M_{k+1}(x,y) + N_{k+1}(x,y)$$

where S(x,y), M(x,y) and N(x,y) are the moving background, target and noise signals. The difference image equation is $D_k(x,y) = F_{k+1(x,y)} - F_k(x,y)$. $[M_{k+1}(x,y) - M_k(x,y)]$ is the covered and uncovered area by the object. $[S_{k+1}(x,y) - S_k(x,y)]$ is the false motion given by the background changes, is filter out by window prediction and $[N_{k+1}(x,y) - N_k(x,y)]$ is the motion by noise, is filter out by preprocessing. The difference image is thresholded with a suitable value. The difference image $D_{tk}(x,y)$ is '1' for $D_k(x,y)>T_h$ (threshold value) and is '0' otherwise. The clusters of nonzero pixels in the difference image indicate motion areas in the image frames. These motion areas are analyzed to detect probable targets.

The Fig. 2 depicts the detail functions carried out in Tracking State. The edge detection, segmentation, template upgradation, window processing, window prediction are carried out in this state. The edge detection is done by applying Sobel's horizontal and vertical edge operator in sequence and resultant image is calculated as follows

$$D(x,y)=\sqrt{(((D_{HorEdge}(x,y))^2+(D_{VertEdge}(x,y))^2)}$$

where $D_{HorEdge}(x,y)$ and $D_{VertEdge}(x,y)$ are images after application of horizontal and vertical edge operator respectively. Then target is segmented and bounding box for the template is computed.

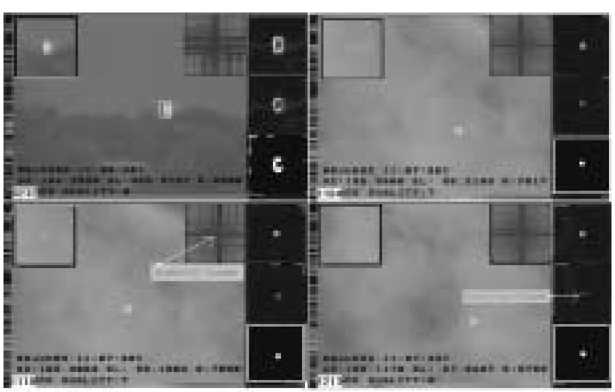


Fig. 3(a)-(d). Results of video tracking system on image sequence

An alternative template is also computed on application of Modified HDT on a small region around the previously (Auto/Manual) located target. The Hausdroff measure [9] from M to I, where M and I are point sets, is h(M,I) = max min ||m-i|| where m ε M, i ε I and ||.|| is any norm. This yields the maximum distance of a point in a set M from its nearest point in set I. The distance transform measures the distance of each pixel in the image from an edge pixel and can be computed efficiently using a two-pass algoritm. The HDT is computed to improve the quality of match between target object and detected target template obtained from previous image frame.

The recovery state is responsible for auto/manual detection and localization of the moving target in video sequences. The automatic detection of the target is achieved by deterministic template matching where template is generated from the detected object in the every previous image frame. Use of template upgradation approach makes the system adaptive to different kind of sequences and avoids any model scale selection.

5 Deterministic Template Matching

The most deterministic way to look for an object in an image is by template matching. Here a template, i.e., an image containing the object only, is matched it to each part of the subsequent frame image. This is called deterministic template matching, and is very efficient when the exact appearance of the object is known. Generally the object may be transformed by rotation, or the exact appearance of the object even without transformations may be unknown. So the normalized template matching will be used to handle transformed objects. The standard detection method in image processing is normalized template matching or matched filtering, where a template $T(x,y)$, $0<=x<x_T, 0<=y<y_T$ is correlated with a part of the image $I(x,y)$. The normalized correlation coefficient $\rho(x_0,y_0)$ will be used as a measure of how well the template fits the image at the position according to

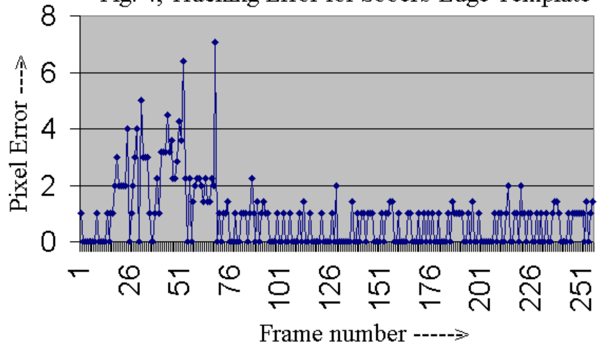
$$\rho(x_0,y_0) = \frac{\Sigma_{x,y}\, T(x,y)\, I_T^{x0,y0}(x,y) - \mu_T\mu_I}{\sigma_T\sigma_I}$$

where

$$I_T^{x0,y0}(x,y) = I(x+x_0,y+x_0)$$
$$0<=x<x_T, 0<=y<y_T$$

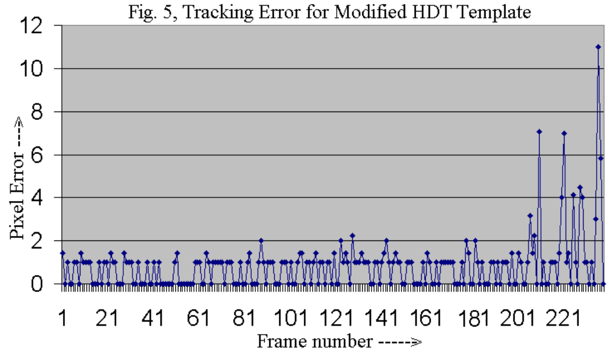
μ_T, μ_I, σ_T and σ_I are the means and variances of and respectively. The summation should go over all nonzero pixels in *T*. ρ is a measure of how well the image part can be approximated by the template T, and in case of a perfect match |ρ|=1. The object is found by looking for a large value of ρ. This approach is optimal to detect a known (deterministic) object in an image disturbed by additive white noise, but in detection applications there are two major drawbacks:

Fig. 4, Tracking Error for Sobel's Edge Template



First, if the object has been transformed in some way, for example, scaled in size, rotated or reshaped, the matched filter may not find the object. In that case, the same transformation has first to be applied to the matched filter. If the parameters of the transformation, e.g., the size of the object, are unknown, several versions of the template has to be created and matched to the image. In this way, the computational complexity will increase heavily.

Fig. 5, Tracking Error for Modified HDT Template



Second, if the exact (un-transformed) appearance of the object is unknown, the matched filter cannot be constructed.

6 Results

The system was tested on real image sequences. The tracking performance was satisfactory. Tracking results on one sequence are shown in Fig 3, where the large rectangular clip represents the processing window and the white rectangle represents the target location. The intermediate processing windows obtained from Sobel's Edge and HDT computation are shown seperately.

The intermediate processing results and effect of various tuning parameters are also shown. The video tracking system is implemented for optical image sequences. The image sequences selected for testing consist of real world environment such as occlusion, shadow, cloud cover, haze, seasonal variation, clutter, moving background and various other form of image degradation. Fig 3(a)-(d) show results obtained on optical image sequence. In these sequences an evident vertical and horizontal camera drift is present. Fig 3(a) shows the accurate detection of the target in tracking state in presence of moving background as tree lines move from one frame to other and slight movement of the target. Fig 3(c) and (d) shows that though target is virtually invisible, system is able to detect it accurately.

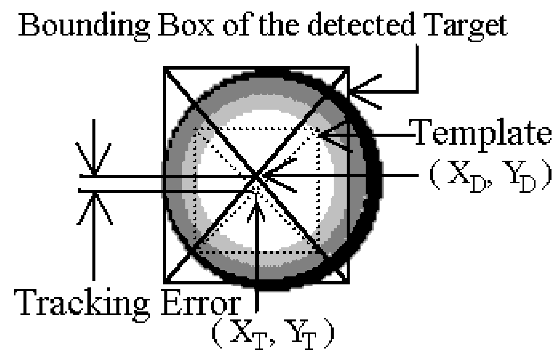


Fig. 6, Estimation of Tracking Error

The tracking error magnitudes for image sequence are shown in Fig.4 and Fig.5. The Tracking Error (ErrT) is defined as the distance between the center of the bounding box of the detected target and template center as shown in the Fig 6. Thus the tracking error is estimated as ErrT = (((XT-XD)2+(YT-YD)2) where (XT,YT) and (XD,YD) are center of the template and detected target respectively. This template is created from the same frame after automatic detection of the target for target tracking in subsequent frame. The tracking error appears to be varying significantly with the target size (approx. 20%) in Sobel's edge based template matching. Here the error magnitude is reduced considerably (within 2 pixels) with the reduction of the target size from 35x35 pixels to 10x10 pixels. On the otherhand, error magnitude in cosiderably (within 2 pixels) reduced and does not vary with the target size in HDT based template matching as shown in Fig. 5. The large error magnitude at the end of the sequence shown in the Fig.5 indicates that system is not able to detect the target due to noisy background where terget is allmost invisible.

7 Conclusion

A complete real time video tracking system for detection and tracking of fast moving target has been described in this paper. Real world data, highlighting the difficulties of target recognition in practical situation are used to demonstrate the effectiveness of our proposed approach. This system achieves automatic detection and also provides semiautomatic and manual detection facilities. The system is tested for optical image sequences. The results confirm that the system performing reliable and accurate tracking for several image sequences.

The proposed technique involves threshold parameters, which are empirically set during testing of system. Further research is planned towards the automatic determination of the thresholds used in this technique. The intelligent parallel hierarchical search strategy may be implemented for reducing the search time during deterministic template matching. Edge also may be precisely extracted by finding the zero-crossing from second derivative of the smoothed image. Hence the tracking error will be reduced significantly with little increase in tracking time. In modified HDT template matching technique, redeced target detection error may be obtained by incorporating efficient search strategy by considering orientation of the target.

8 References

[1] A.Neri, S. Colonnese, G. Russo, P. Talone, Automatic moving object and background separation, Signal Processing 66 (1998), pp. 219-232.

[2] A. L. Gilbert, M. K. Giles, G. M. Flachs, R. B. Rogers and Yee Hsun U, “A Real-Time Video Tracking System”, IEEE Trans. On PAMI, Vol. 2, No. 1, 1980, pp. 47-56.

[3] Roland Mech, Michael Wollborn, A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera, Signal Processing 66 (1998), pp. 203-217.

[4] B. K. P. Horn and B. G. Schunck, “Determining Optical Flow”, Artificial Intelligence, 17(3), 1981, pp. 185-203.

[5] Kristine E. Matthews, Nader M. Namazi, A Bayes Decision Test For Detecting Uncovered Background and Moving Pixels in Image Sequences, IEEE Transaction On Image Processing Vol.7, No.5, May1998, pp.720-728.

[6] James Arnold, Scott Shaw Henry Pasternack, Efficient Target tracking Using Dynamic Programming, IEEE Transaction On Aerospace And Electronic Systems, Vol.29, No.1, Jan 1993, pp.44-56.

[7] N.C. Mohanty, Computer Tracking Of Moving Point Targets In Space, IEEE Transaction On Pattern Analysis And Machine Intelligence, Vol.PAMI-3, NO.5, Sept 1981,pp. 606-610.

[8] A. Reid, V Markandy and S. Wang, “Motion estimation for moving target detection, IEEE Transaction On Aerospace and Electronic System, Vol. AES 32, No.2.July 1996, pp 866-872

[9] Clark F. Olson and Daniel P. Huttenlcher, Autometic Target Recognition by Matching Oriented Edge Pixels, IEEE Transaction on Image Processing, Vol.6, No.1, January 1997