

# Small Object Detection and Tracking: Algorithm, Analysis and Application

U.B. Desai, S.N. Merchant, Mukesh Zaveri, G. Ajishna,  
Manoj Purohit, and H.S. Phanish

SPANN Lab, Electrical Engineering Dept., IIT Bombay - 400076  
ubdesai@ee.iitb.ac.in

**Abstract.** In this paper, we present an algorithm for detection and tracking of small objects, like a ping pong ball or a cricket ball in sports video sequences. It can also detect and track airborne targets in an infrared image sequence. The proposed method uses motion as the primary cue for detection. The detected object is tracked using the multiple filter bank approach. Our method is capable of detecting objects of low contrast and negligible texture content. Moreover, the algorithm also detects point targets. The algorithm has been evaluated using large number of different video clips and the performance is analysed.

## 1 Introduction

Detection and tracking of small and fast moving objects is increasingly becoming important for the sports industry. Critical decisions are now based on the exploitation of such algorithms. Decision based on very slow motion replays in football and cricket is already highly pervasive. The objective of this paper is to develop algorithms which can further assist in making such decisions as well as in analyzing sports video sequences. The paper focuses on a small segment of this large task, namely, detection and tracking of small objects in sports video sequences.

Another important application of detection and tracking is in IR (infrared) images where, the target of interest could be either a point target or an approaching one (varying target size). This problem is challenging as the intensity of the target and that of the cluttered background varies with time. Even the shape and size of the clouds and target may vary from one time instant to another. Because of these reasons, the methods based on spatial processing, [1] and optical flow [2] do not help.

In literature, various approaches have been proposed for object detection and tracking. Techniques such as background registration [3] or background extraction [4] may not be feasible in sports video sequence due to the non-stationary nature of the background. Detection of a small moving object based on segmentation of each frame [5] will not be viable, in our case, due to low contrast image and lack of any texture information. In this paper we propose a method to detect small and fast moving object in sports video sequences using motion as the primary cue. This work is based on our earlier work on point target detection [6].

Along with detection, tracking algorithm also plays an important role. A variety of methods for tracking multiple targets based on multiple hypothesis tracking (MHT) and joint probability data association filter (JPDA) have been proposed in literature [7,8]. A dynamic programming approach (DPA) has been proposed in [9], and the truncated sequential probability ratio test (SPRT) is presented in [10] to detect and track targets with straight line trajectories. All these methods are computationally expensive and have little scope for real time application. Moreover, these methods are able to track only linear trajectory.

We track the targets using two different methods; (i) modified pipeline algorithm [11] and (ii) filter bank approach [12,13]. The key feature of the filter bank approach is its ability to track maneuvering as well as non-maneuvering trajectories in the absence of any a priori information about the target dynamics.

## 2 Object Detection

### 2.1 Wavelet Based Detection

We use wavelet transform for temporal filtering [6] as it enables one to detect and characterize the dynamical behavior of elements present in the scene. We have used the Haar wavelet because with increase in the number of wavelet filter coefficients, larger number of image frames are needed by the transform to do a proper job of detection. The temporal multiscale decomposition facilitates the construction of the intensity change maps which indicate whether there is temporal change or not. A two-hypotheses likelihood ratio test is then applied to validate the temporal changes at each scale. By exploiting this likelihood ratio test, the issue of motion detection is solved in statistical frame work.

In order to make the detection scheme robust to clutter and noise, post processing is incorporated. As stated above temporal multiscale decomposition and hypothesis testing provide the (binary) change detection map. In this map, nearby small regions are merged using binary morphological operation, namely, closing (dilation followed by erosion). Next, the change detection map is segmented and all segments having a size larger than a pre-defined threshold are removed. Small size clutter which appear like small targets are eliminated by comparing local contrast with a predefined threshold. If it crosses the threshold it will be a moving target.

### 2.2 Gradient Based Detection of Small Targets

It is often difficult to detect small objects of low contrast in a highly evolving background. This algorithm[14], based on spatial gradient thresholding and region merging, is proposed to overcome the problem and effectively detect small objects of low contrast in evolving clouds.

The algorithm uses image-differencing technique to register motion. The regions corresponding to variation between frames, wherein there is a target motion or any displacement would be captured in the difference image. Also, this eliminates any stationary clutter, assuming the imaging facility is stationary.

The spatial gradient of the difference image is then obtained by applying the Sobel operator which identifies the edges in the difference image. The edges of a target are usually well defined and stronger than those of clouds. Hence, the target motion registered in the difference image can be separated from the temporal differences registered by moving (or evolving) clouds, using this as a cue. A gradient magnitude thresholding is done to highlight target regions i.e., the regions of interest (ROI) among the low magnitude variations contributed by clouds.

The processed images however contain stray pixels corresponding to isolated noise pixels as well as cloud edge variations. These pixels can be eliminated and the regions of interest can be located by using a region merging technique like binary morphological closing. Once the ROI are located in the original frame, local contrast of these regions within specific windows are inspected. The pixel is then declared as a target pixel if the local contrast exceeds a predefined threshold.

### 3 Object Tracking

For object tracking we proposed two algorithms, first one deals with linear and slow maneuvering target. And the second algorithm is able to track arbitrary trajectories of the objects.

#### 3.1 Modified Pipeline Based Tracking

The pipeline algorithm is able to detect and track targets having 1-2 pixel movement per frame. To track a target with large movement ( $\pm 20$  pixels per frame), the window size needs to be increased and hence it leads to false alarm due to increased search neighborhood in continuity filtering algorithm (CFA). Consequently, taking centroid position of Temporal Window Column (TWC) as a detected target position is not correct, as number of additions per pixel increase by  $O(N^2)$  with increase in maximum detectable target movement, where N is the number of pixels in a frame. To overcome these limitations, modified pipeline algorithm is developed, a variable slope pipe shown in Figure 1 is used. Pipeline

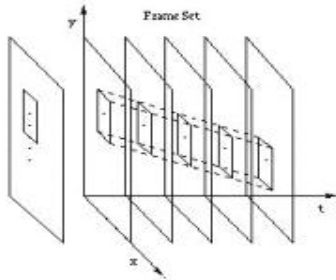


Fig. 1. Modified Test Pipe

algorithm consists of two major components: (i) fixed length TEST PIPE of temporally adjacent image frames. At each iteration a frame from the bottom of the test pipe is discarded and new frame is added to the top of the pipe and other frames are shifted in the pipe by one position. (ii) AND PIPE (AP), consists of two image frames and a single blank frame called Target Frame (TF). The CFA uses the AND PIPE to test continuity of the target pixel. The algorithm is as follows: (1) Initialize the AND PIPE (AP) by adding a frame to the top at each cycle. (2) Initialize the TEST PIPE (TP) with  $n - 1$  frame cycles, where  $n$  is the number of frames in TP. (3) At each time step, (a) A constant acceleration based Kalman filter is used in a predictor mode, which predicts the target position in the next frame. (i) Update the AP (ii) Apply the CFA in AP. (b) (i) Update the TP by adding the output frame of the AP in each cycle to the top of the TP. (ii) Form a search window around the predicted position given by the Kalman filter. Apply the TWC at each pixel  $(X_i, Y_i)$  in the search window and sum the intensities. This greatly reduces the computational load. This sum is given by:

$$S(X_i, Y_i) = \sum_{k=1}^n \sum_{x=-w/2}^{w/2} \sum_{y=-l/2}^{l/2} I(X_i + x, Y_i + y, k)$$

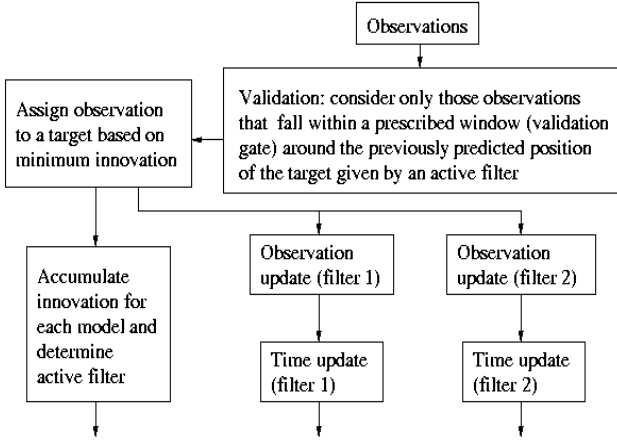
where  $n, w$  and  $l$  are the dimension of the TWC in the TP. (iii) If sum is greater than threshold (determined by possible target intensities) then go to Step 3(d), else consider it as occlusion and go back to Step 3(a) for the next time step. (d) Compute the intensity centroid. Find the Euclidean distance between the current position and the previous position and compare it against a threshold. If it is below the threshold then, accept as trajectory pixel and record in the TF else reject.

The modified pipeline algorithm requires coarse estimate of initial velocity, to overcome this, the output of target detection phase is used in following manner: First, the candidate target list is formed using the first two frames, and secondly using nearest neighbour technique with maximum velocity constraint, candidate is associated from one list to another list, which provides estimation of velocity and a new target track is initiated if association is not possible for any candidate target in the list.

### 3.2 Multiple Filter Bank Approach

Using a single tuned filter, it is difficult to track complex maneuvering trajectories. We propose a method to track multiple point target movement using multiple filter bank [12]. The filter bank consists of different types of filters. For example, in a bank of two filters, one could be a constant velocity filter and the other could be based on a maneuver model.

In the filter bank, all the filters run in parallel. Nevertheless, at any given time, only one filter is designated as an active filter - this is the filter whose output is selected as the predicted position of the target. The validation gate is formed around this predicted position and only validated observations (i.e. observation falling inside this validation gate) are used for data association. In



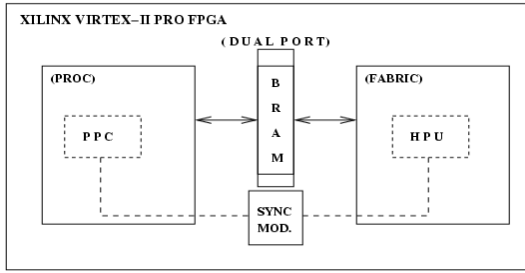
**Fig. 2.** Multiple Filter Bank algorithm with two filters

our case, Munkres’ optimal data association algorithm is used for data association based on nearest neighbor method. The state vector of other models in a filter bank is also updated using the same observation. The decision on which filter becomes the active filter is based on minimum average innovation. The logical steps for multiple filter bank algorithm are described in Figure 2. An approach based on the use of multiple filters has been explored earlier [15]. But in the proposed method switch-over between the filters in the bank is based on single-step decision logic, and consequently, is computationally more efficient and performance-wise more robust. For filter switch-over, the innovation error is accumulated over the past iterations for each filter in the filter bank. It is averaged and compared with that of other filters. Based on minimum averaged innovation error for the filter, the switch-over takes place. From our simulations, we observe that a filter bank with two filters, one based on constant acceleration model and the other based on acceleration being modeled as colored noise, is able to track both non-maneuvering and maneuvering targets.

### 3.3 FPGA Based Tracking Hardware Implementation

Xilinx Virtex-II Pro FPGAs provide one or more embedded PowerPC processor cores along with FPGA fabric. Combining a PowerPC processor core with the Virtex-II FPGA fabric brings a huge amount of flexibility, the processor and FPGA are used together in a way that makes the best use of the power of each element. The PowerPC provides many different interfaces that are designed for use in different ways, which are used to enhance the overall performance of the system and to achieve right balance between the workload of the PowerPC and FPGA. FPGA logic can be connected to PowerPC interfaces, with software for the CPU interacting with HDL written for the FPGA.

The Hardware acceleration block takes the Image pixels processes them and writes them back and thus achieving result better than processor in terms of

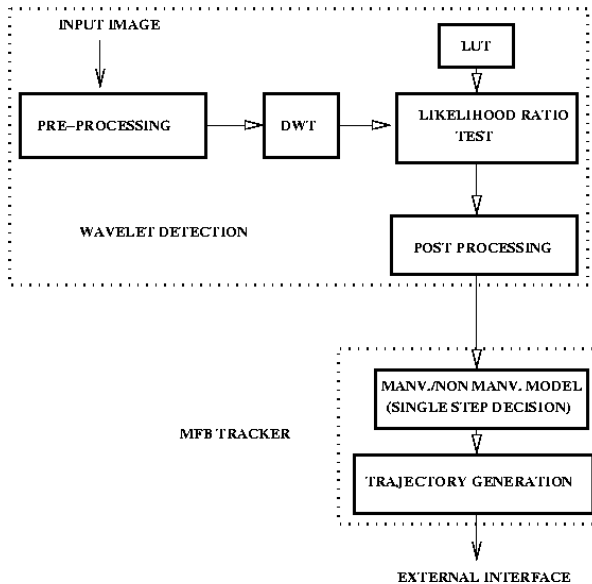


**Fig. 3.** Virtex-II pro Based Embedded system

speed. Function of the synchronisation module is to control data access to the Block RAM and is used for data synchronisation between PowerPC and FPGA based hardware processing unit. Architecture based decoupling (Figure 3) is possible as Block RAMs are dual ported, with one side connected to the processor and other to logic.

PowerPC feeds the hardware acceleration block with data and also controls and monitors the operation of that logic. This provides a hardware framework that can be used to implement many other different types of FPGA function block (wavelet detection, etc.).

Thus the main features of Virtex-II pro (platform FPGA) based implementation is that both PowerPC and logic used for design, use of RTOS (e.g. Xilkernel) for embedded processor, use of customized logic (hardware acceleration unit) for time critical part alongwith synchronisation unit for data synchronisation between processor and hardware acceleration unit.



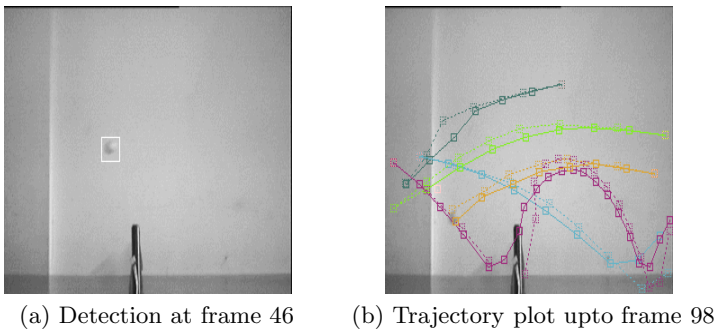
**Fig. 4.** Algorithm block diagram

## 4 Simulation Results and Performance Analysis

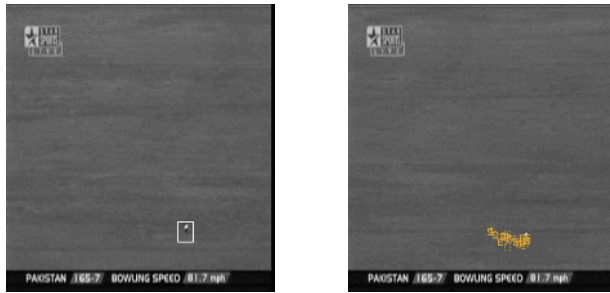
In our simulations, we have used two filters: constant acceleration (CA) and Singers' maneuver model (SMM) for the multiple filter bank. For the simulations, the tracker is setup after object continuity is found in three consecutive frames in the sequence. We have evaluated the performance of the proposed algorithm using a number of sequences. Simulation results are presented for two sport sequences; cricket and ping pong.

The ping pong sequence has 100 frames and was captured with a stationary camera. In the sequence, the ball does not appear as a full circle due to low contrast and shadow effects. Our proposed detection algorithm is able to detect and track the ping pong ball with such low contrast. Figures 5-(a) and 5-(b) represent the output of wavelet based detection algorithm at frame 46 and tracking of the ping pong throughout the sequence. The ping ball enters and leaves the sequence number of times. To discriminate entering and leaving time instant, each trace of the tracked ball is represented using different colors. In Figures 5-(b) and 6-(b) the real trajectory is shown with a solid line, whereas the predicted trajectory is shown using a dotted line with the same color. In the cricket video sequence we have 32 frames for detection and tracking. In the cricket sequence, unlike the ping pong ball sequence, the background has significant motion. Another challenging aspect is the varying size of the ball due to zoom in and zoom out of the camera. Our algorithm is able to detect and track the cricket ball effectively in spite of these challenges. The detection output for the cricket sequence is shown in Figure 6-(a). Figure 6-(b) depicts the tracked trajectory of the cricket ball upto frame 15.

We used our algorithms to detect and track approaching targets in IR sequence also. Simulation results for two clips are depicted here. (Note, due to space limitations, we have depicted the details of trajectories in clip 1 only.) For clip 1, for all trajectories, initial target signature is of  $3 \times 3$ , gradually increasing to a maximum size of  $15 \times 11$ . The enlarged view of the target signature is shown in Figure 7(c). First trajectory is generated using constant velocity, coordinated turn and constant acceleration model. The initial position and initial velocity are set to  $(9km, 7km)$  and  $(280.0m/s, 280.0m/s)$ . Trajectory takes total of four

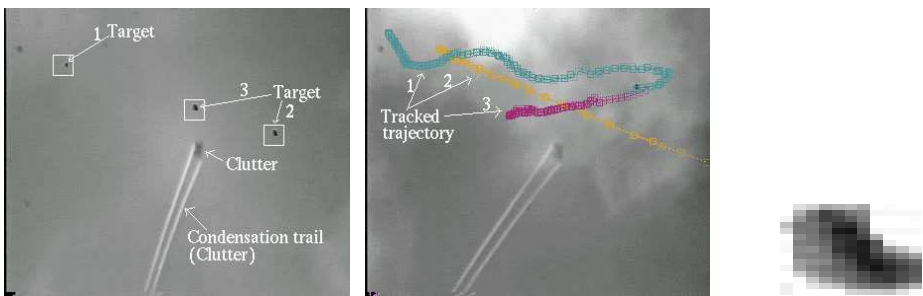


**Fig. 5.** Ping Pong Ball Detection and Tracking



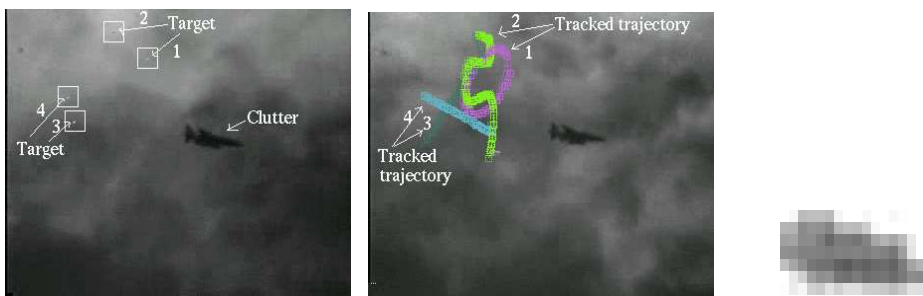
(a) Detection at frame 26 (b) Trajectory plot upto frame 15

**Fig. 6.** Cricket Ball Detection and Tracking



(a) Detection at frame 20 (b) Trajectory Plot upto frame 69 (c) IR target signature (Enlarged view)

**Fig. 7.** Approaching Target Detection and Tracking (clip 1)

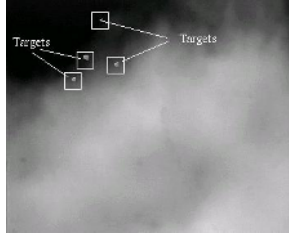


(a) Detection at frame 4 (b) Trajectory Plot upto frame 69 (c) IR target signature (Enlarged view)

**Fig. 8.** Approaching Target Detection and Tracking (clip 2)

turns: (a) from frame 15 to 22 with turn angle  $15^\circ$ , it results in  $\approx 10.5g$  acceleration, (b) from frame 30 to 36 with turn angle  $-15^\circ$  ( $\approx 10.27g$  acceleration), (c) from frame 43 to 52 with  $12^\circ$  ( $\approx 8.8g$  acceleration) and (d) from frame 58 to 67 with  $-12^\circ$  ( $\approx 6.1g$  acceleration). For projecting the trajectory on to the image plane, the trajectory plane is assumed at 8km depth and rotated about (X,Y,Z) axis by  $(15^\circ, 25^\circ, 12^\circ)$  respectively. Second trajectory is generated with initial





**Fig. 9.** Detected and marked targets

position  $(7km, 7km)$ , initial velocity  $(300m/s, 270m/s)$  and initial acceleration  $(100.0m/s^2, 80.0m/s^2)$ . The trajectory plane is assumed to be at 7km and it is rotated by  $10^\circ$  about X-axis and by  $40^\circ$  about Y-axis for projection on to the image plane. Finally, the third trajectory is generated with initial position  $(10km, 7km)$  and initial velocity  $(320m/s, -260m/s)$ . For projecting the trajectory on to the image plane, the trajectory plane is rotated by  $20^\circ$  about X-axis and by  $30^\circ$  about Y-axis. Figures 7(a) and 7(b) represent the detection output at frame 20 and the tracked trajectory upto frame 69. The condensation trail formed at the wing tips appears as clutter in the sequence. Clip 2 is similar to clip 1 except for: (a) it has 4 targets, (b) maximum target size is  $16 \times 7$ . Figures 8(a) and 8(b) depict the detected target at frame 4, and the tracked trajectories. The enlarged view of the target signature is shown in Figure 8(c). The gradient based detection algorithm was used to detect targets in an IR clip. The sequence consisting of four targets embedded in evolving cloud clutter, were detected are shown in Figure 9.

## References

1. Braga-Neto, U., Choudhary, M., Goutsias, J.: Automatic target detection and tracking in forward-looking infrared sequences using morphological connected operators. *Journal of Electronic Imaging* ((In press) 2004)
2. Vaswani, N., Agrawal, A.K., Zheng, Q., Chellappa, R.: Moving object detection and compression in IR sequences. In Bhanu, B., Pavlidis, I., eds.: *Computer Vision beyond the Visible Spectrum*. Springer (2003) 153–177
3. Chien, S.Y., et al.: Efficient moving object segmentation algorithm using background registration technique. *IEEE Transactions on Circuits and Systems for Video Technology* **12** (2002) 577–586
4. Durucan, E., Ebrahimi, T.: Change detection and background extraction by linear algebra. *Proceedings of IEEE* **89** (2001) 1368–1381
5. Tsaig, Y., Averbuch, A.: Automatic segmentation of moving objects in video sequences: A region labeling approach. *IEEE Transactions on Circuits and Systems for Video Technology* **12** (2002) 597–612
6. Mukesh A. Zaveri , S.N. Merchant and Uday B. Desai: Multiple single pixel dim target detection in infrared image sequence. In: *Proc. IEEE International Symposium on Circuits and Systems, Bangkok* (2003) 380–383
7. S. Blackman, R. Dempster and T. Broida: Multiple Hypothesis Track Confirmation for Infrared Surveillance Systems. *IEEE Transactions on Aerospace and Electronic Systems* **29** (1993) 810–824

8. Y. Bar-shalom and T. E. Fortmann: Tracking and Data Association. Academic Press (1989)
9. Yari Barniv: Dynamic Programming Solution for Detecting Dim Moving Targets. IEEE Transactions on Aerospace and Electronic Systems **21** (1985) 144–156
10. S. Blostein and T. Huang: Detecting small, moving objects in image sequences using sequential hypothesis testing. IEEE Transactions on Signal Processing **39** (1991) 1611–1629
11. Mukesh A. Zaveri, Anant Malewar, S.N. Merchant and Uday B. Desai: Wavelet Based Detection and Modified Pipeline Algorithm for Multiple Point Targets Tracking in InfraRed image sequences. In: Proceedings of 3rd Conference ICVGIP 2002, Ahmedabad, India (2002) 67–72
12. Mukesh A. Zaveri, Uday B. Desai and S.N. Merchant: Tracking multiple maneuvering point targets using multiple filter bank in infrared image sequence. In: Proc. IEEE International Conference on Acoustics, Speech, & Signal Processing, Hongkong (2003) 409–412
13. Uday B. Desai, S. N. Merchant, Mukesh A. Zaveri: Detection and Tracking of Point Targets. In: Proceedings of the 5th International Conference on Advances in Pattern Recognition (ICAPR - 2003) Invited Paper (to appear), Calcutta, India (2003)
14. Ajishna G.: Target detection in ir sequences. Master's thesis, Indian Institute of Technology Bombay (June 2005)
15. James R. Cloutier, Ching-Fang Lin and Chun Yang: Enhanced Variable Dimension Filter for Maneuvering Target Tracking. IEEE Transactions on Aerospace and Electronic Systems **29** (1993) 786–797