Small and fast moving object detection and tracking in Sports Video Sequences

Mukesh A. Zaveri S. N. Merchant Uday B. Desai
SPANN Lab, Electrical Engineering Dept., IIT Bombay - 400076.
Email address : [mazaveri,merchant,ubdesai]@ee.iitb.ac.in

Abstract

We propose an algorithm for detection and tracking of small and fast moving objects, like a ping pong ball or a cricket ball, in sports video sequences. For detection, the proposed method uses only motion as a cue; moreover it does not use any texture information. Our method is able to detect the object with very low contrast and negligible texture content. Along with detection, we also propose a tracking algorithm using the multiple filter bank approach. Thus we provide a complete solution. The tracking algorithm is able to track maneuvering as well as non-maneuvering movements of the object without using any apriori information about the target dynamics.

1. Introduction

Detection and tracking of small and fast moving objects is increasingly becoming very 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 the paper is to develop algorithms which can further assist in making such decision as well 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. In literature, various approaches have been proposed for object detection and tracking. In [7] a method is proposed on background registration; but for sports video sequence, background registration [7] or background extraction [2] may not be feasible due to the non-stationary nature of the background. Detection of a small moving object based on segmentation of each frame [8] will not be viable, in our case, due to low contrast image and lack of any texture information. A nonparametric approach has been proposed in [1], but estimation of kernel density is time consuming.

In this paper we propose a method to detect small and fast moving object in sports video sequences using motion as the only cue. This work is based on our earlier work on point target detection and tracking [3]-[4]. Our approach exploits the wavelet transform for motion detection. The main thrust of the proposed approach is to detect the object under low contrast, without using any texture information or any other feature other than motion. After detection, we track the object using a filter bank approach proposed by us in [4]. The key feature of the filter bank approach is its ability to track maneuvering as well as non-maneuvering movements of the object in the absence of any apriori information about the object dynamics. In the current work targets of size of a few pixels can be detected and tracked.

2. Wavelet Based Small Moving Object Detection

We use wavelet transform for temporal filtering [3] because the wavelet transform enables one to detect and characterize the dynamical behavior of the elements present in the scene. In the absence of any texture information, to detect small moving objects in cluttered background, a longer temporal integration is required. We have used the Haar wavelet because, with an increase in the number of wavelet filter coefficients, larger number of image frames are needed by the wavelet transform to do a proper job of detection. This in turn introduces larger delay in making the decision. With the Haar wavelet this delay is minimum. 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 the likelihood ratio test, the motion detection issue is solved in statistical frame work. Two competing hypotheses are compared : hypothesis H_0 (no temporal change at p) and hypothesis H_1 (temporal change at p). The log-likelihood ratio corresponding to hypotheses H_1 and H_0 is derived and the decision step is formalized as:

$$\psi^k(p) \underset{H_1}{\overset{H_0}{\substack{<\\ \geq}}} \lambda$$

where $\psi^k(p)$ is the resulting expression of the log-likelihood ratio at scale k, in the maximum likelihood sense

$$\psi^k(p) = \frac{1}{2\sigma_k^2}\left[\frac{1}{N}\left(\sum_{i=1}^{N} D^k(p_i)\right)^2 + \frac{1}{\sum x_i^2}\left(\sum_{i=1}^{N} x_i D^k(p_i)\right)^2 + \frac{1}{\sum y_i^2}\left(\sum_{i=1}^{N} y_i D^k(p_i)\right)^2\right]$$

and follows a χ^2 distribution with three degrees of freedom. λ is a threshold which may be inferred from tables of statistical laws, N is the size of the window in terms of pixels centred at point p_i and (x_i, y_i) indicates the relative location of pixels with respect to the centre of the window. σ_k^2 is the variance of the pixel intensity within the window.

2.1. Clutter Removal

In order to make the detection scheme robust to clutter and noise, post processing is incorporated. In [3] the post processing step is used to detect the point target only. This step needs to be modified for small object detection. To remove isolated points from the change detection map, first binary morphology operation, namely, opening is performed on the change detection map. Next the change detection map is segmented and all segments having a size larger than a threshold defined by δ_{th}, are removed. The edge effects and small size clutter which appear like small targets are eliminated by comparing local contrast $lc(x_n, y_n)$ with a threshold ρ. If it crosses the threshold it will be a small moving object. Here, $lc(x_n, y_n)$ is defined as $lc(x_n, y_n) = \mid I(x_n, y_n) - \frac{1}{s_i} \sum_{(x_m,y_m)\,\epsilon\,N_w} I(x_m, y_m) \mid$ where $I(x_n, y_n)$ is the gray level value of the pixel at (x_n, y_n), s_i is the number of pixels in the neighborhood window centered at (x_n, y_n). N_w represents the neighborhood window. The output of the above step gives a list of detected small moving object that is used by the tracking algorithm. It is treated as a set of observations available at current time instant.

3. Tracking Using Filter Bank

After detecting the small object in the sequence, the next step is to track this object that could be non-maneuvering or maneuvering. Using a single tuned filter, it is difficult to track such trajectories. We use our earlier proposed method to track multiple point target movement using multiple filter bank [4]. 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 proposed method switch-over between the filters in the bank is based on single-step decision logic, and consequently, there is no delay in estimation and provides tracking in real time. The constant acceleration or constant velocity based Kalman filter is able to track non-maneuvering objects. The constant acceleration model performs well when acceleration is in the direction of velocity. It does not work well with highly maneuvering object. Therefore, we add one more filter, a Kalman filter based on Singer's model [5], which is used to track maneuvering objects. In this model, the acceleration is modeled as colored noise [9]. We have used constant acceleration (CA) and Singer's maneuver model (SMM) for our simulations.

The tracking algorithm generates several tracks and we need to identify which track corresponds to the object under consideration. In order to achieve this correspondence, we use the nearest neighbor method for data association. Error measure value (innovation) calculation is done for each observation with respect to every object in the validation gate. This is followed by Munkres' optimal data assignment algorithm [6], which is used to assign an observation to a track. If no measurement is associated with a track for over several consecutive image frames, then the filter bank for that object is eliminated and the track is terminated. Data association is followed by a single step decision logic for filter switch-over.

3.1. Single step decision logic

We use a single step decision logic, which provides a measure to characterize the behavior of the object in the absence of any apriori information. At every time instant, an observation that is assigned by Munkres' algorithm, is given to all the filters in the filter bank. The filters now update their states independent of each other. The innovation error is accumulated over the past iterations for each filter in the filter bank. It is averaged and compared with that of the other filters. The switch-over takes place based on the minimum averaged innovation error for the filter. Let, $\mathbf{z}(k)$ be an observation associated with a particular track by data association method, and $\hat{\mathbf{z}}(k|k-1)$ be the predicted measurement at time k, then the innovation is given by $\tilde{\mathbf{z}}(k) = \mathbf{z}(k) - \hat{\mathbf{z}}(k|k-1)$. Filter switch over from filter i to filter j at time instant k takes place if $e_i(k) > e_j(k)$ where, e_i is the average innovation for the i-th filter at time instant k and is defined as

$$e_i(k) = \frac{1}{s}\sum_{m=k-s+1}^{k} v_i(m)$$

Here, s is size of the sliding window. $v_i(m)$ is the innovation for filter i at time instant m and is defined as

$$v_i(m) = \tilde{\mathbf{z}}^T(m)\mathbf{S}^{-1}(m)\tilde{\mathbf{z}}(m)$$

where S (diagonal matrix) is the innovations covariance matrix. The above steps make it possible to track both maneuvering and non-maneuvering objects.

4. Simulation Results

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. The state parameters for each model used in tracking are initialized using the target position found in the first two frames of the sequence. We have evaluated the performance of the proposed algorithm using a number of sequences. Simulation results are presented only for two sequences; cricket and ping pong (due to space limitations). All video sequences, base video and video with the ball being tracked are available at http://www.geocities.com/svc032003.

The ping pong sequence has 100 frames and was captured with a stationary camera. Moreover, the background remains stationary. The main objective here is to detect and track the ball and discriminate it from shadow effects. 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 1-(a) and 1-(b) represent the output of wavelet based detection algorithm at frame 46 and 47 for the ping pong sequence. Figure 1-(c) depicts tracking of the ping pong ball 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 1-(c) and 2-(c) 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. We wanted to illustrate that our algorithm works with short video sequences too. The key challenge here is to track only the ball and not the movement of the players. Similar to the previous case, the ball has low contrast and suffers from shadow effect. The players are can be segmented using any edge operators that exploits high contrast. 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 detect and track the cricket ball effectively in spite of these challenges. This cricket sequence was captured from cable TV using a frame grabber card from *pixel view* on a Pentium machine. The captured sequence is a mixture of multiple camera output. The original sequence is in color, but for detection and tracking, the sequence is converted into gray value for fast execution of the algorithm. The detection output for the cricket sequence is shown in Figure 2-(a) and 2-(b) for frame 26 and 27. Figure 2-(c) depicts the tracked trajectory of the cricket ball at frame 15. Figure 3-(a) represents the changed detection map for the cricket sequence at frame 28. There are isolated points in the change detection map. The binary morphology operation helps in removing such isolated points from the change detection map. For the cricket ball sequence, filter switch over plot and prediction error in position are depicted in Figures 3-(b) and 3-(c).

5. Conclusion

In the absence of any feature information other than motion, our wavelet based object detection algorithm is able to detect small objects in real video sequences. Our multiple filter bank based tracking algorithm uses only two filters, namely, CA and SMM, for tracking. It successfully tracks maneuvering and non-maneuvering object movements in the presence of low contrast, background movement and movement of players. It does not use any apriori information about the target dynamic.

In future we propose to generalize our work to detect and track multiple objects, including players, for example tracking the ball and the bat in cricket, or tracking the foot and the ball in case of soccer.

Acknowledgments

The authors would like to acknowledge Dipan Mehta and Deepak Nayak of Dizzy Research Co. for providing the various sports video sequences.

References

[1] Ahmed Elgammal et.al. Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance. *Proceedings of IEEE*, 90(7):1151–1163, July 2002.

[2] E. Durucan and T. Ebrahimi. Change Detection and Background Extraction by Linear Algebra. *Proceedings of IEEE*, 89(10):1368–1381, October 2001.

[3] M. A. Zaveri, S.N. Merchant, U. B. Desai. Multiple Single Pixel Dim Target Detection in InfraRed Image Sequence. In *Proceedings of IEEE International Symposium on Circuits and Systems*, pages 380–383, Bangkok, May 2003.

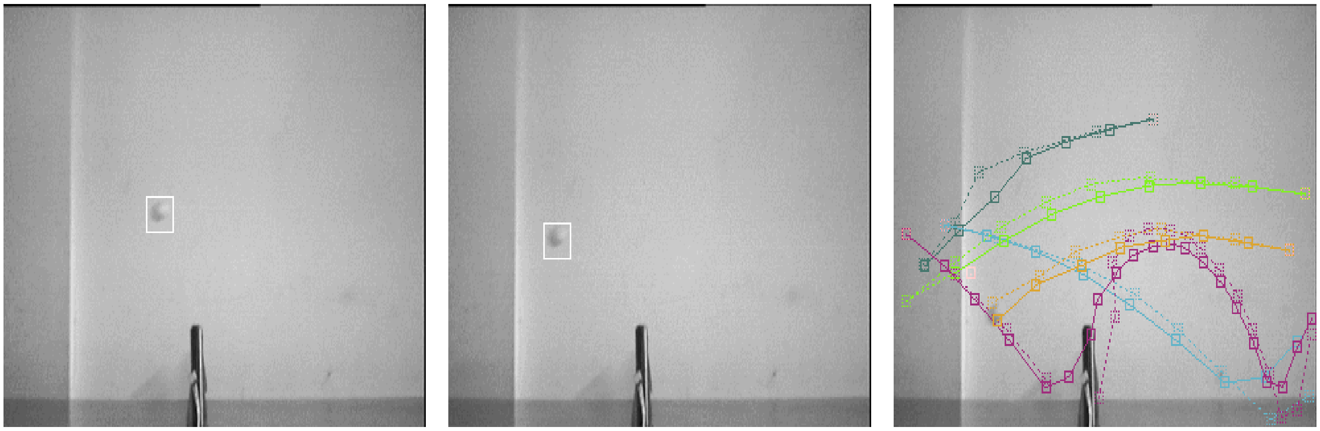
[4] M. A. Zaveri, U. B. Desai and S.N. Merchant. Tracking Multiple Maneuvering Point Targets using Multiple Filter Bank in InfraRed Image sequence. In *Proceedings of IEEE International Conference on Multimedia & Expo (ICME) 2003*, pages 369–372, Baltimore, July 2003.

[5] Robert A. Singer. Estimating Optimal Tracking Filter Performance for Manned Maneuvering Targets. *IEEE Transactions on Aerospace and Electronic Systems*, 6(4):473–483, July 1970.

[6] Samuel S. Blackman. *Multiple-Target Tracking with Radar Applications*. Artech House, Inc., Boston, 1986.

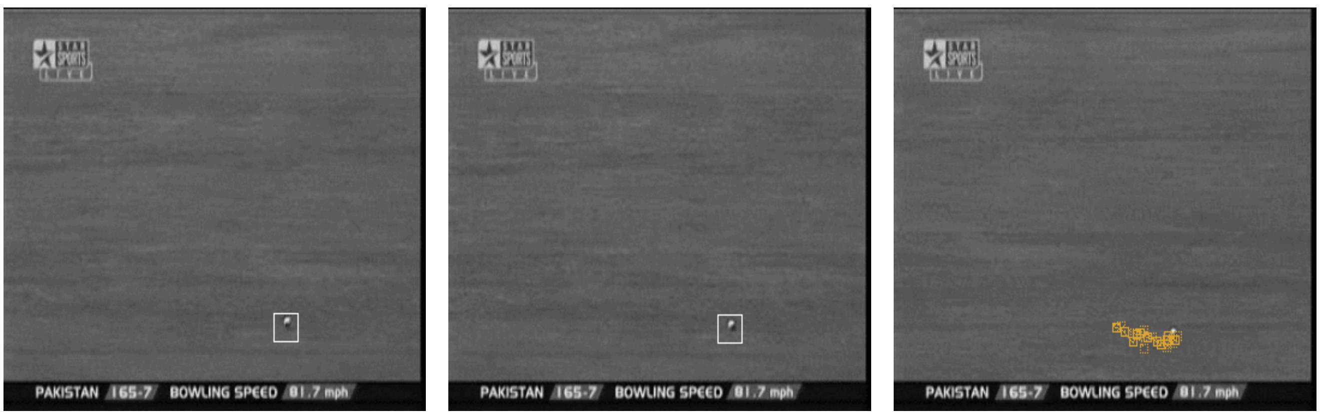
[7] Shao-Yi Chien et.al. Efficient Moving Object Segmentation Algorithm Using Background Registration Technique. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(7):577–586, July 2002.

[8] Y. Tsaig and A. Averbuch. Automatic Segmentation of Moving Objects in Video Sequences: A Region Labeling Approach. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(7):597–612, July 2002.



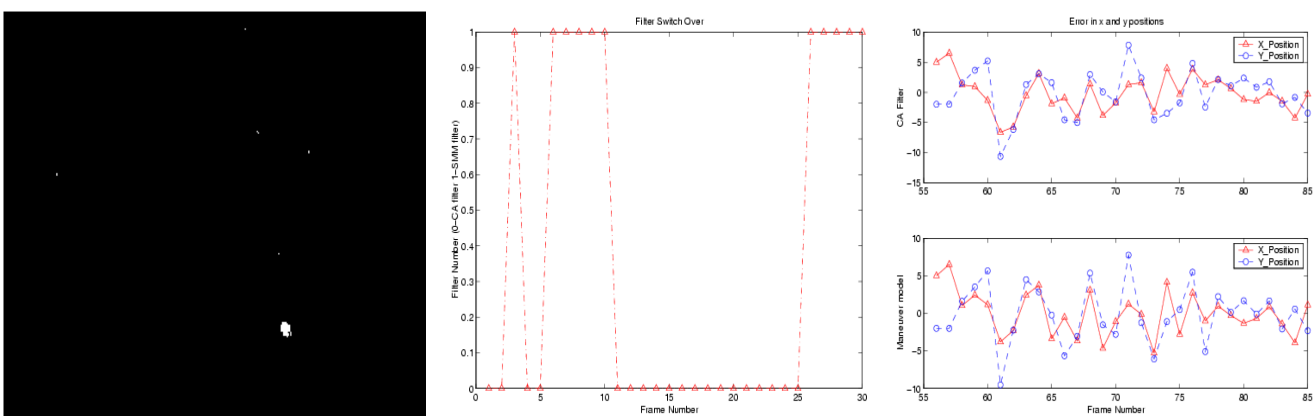
(a) at frame 46 (b) at frame 47 (c) trajectory plot at frame 98

Figure 1. Ping Pong Ball Detection and Tracking



(a) at frame 26 (b) at frame 27 (c) trajectory plot at frame 15

Figure 2. Cricket Ball Detection and Tracking



(a) Change Detection Map at frame 28 (b) Filter Switch Over Plot (c) Prediction Error Plot

Figure 3. for Cricket Ball Tracking

[9] Y. Bar-shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, 1989.