DETECTING AND TRACKING MOVING OBJECTS IN SEQUENCES OF COLOR IMAGES

Hugh L. Kennedy

Pacific Noise and Vibration (PNV) Pty. Ltd.
PO Box 429, Canberra ACT 2601, Australia, hkennedy@pnv.com.au.

ABSTRACT

A statistical change detector, implemented as a zero-latency finite-memory filter, is used to identify anomalies in temporal pixel statistics. An F-distributed test statistic is computed for each pixel and used in a hypothesis test. The tracker, with automatic track initiation and termination, uses a low-complexity pairwise Joint Probabilistic Data Association (JPDA) algorithm, which has been restricted to consider clusters (sub-problems) containing no more than two tracks. The track state and clutter model are augmented to include color. The detector and tracker are used to process sample video data.

Index Terms – Object detection, Image motion analysis, Tracking filters, Surveillance.

1. INTRODUCTION

Automated surveillance systems are typically composed of signal processing and data processing subsystems [1]. The function of peak detection is usually executed within the former subsystem, while the function of tracking, the latter. In many systems the distinction between these functions is not clear. Tight coupling or merging of the functions can yield improved performance; however, development and maintenance effort is increased due to increased system complexity.

The review of 'classical' multi-target tracking methods presented in [2] represents one end of the design spectrum, where only the coordinates of the peak maxima, in a continuous measurement space are passed from the signal processing subsystem to the data processing subsystem. Track-Before-Detect (TBD) techniques [3,4] represent the other end of the spectrum where tentative tracks are processed in the discrete (digital) measurement space prior to confirmation (i.e. when the track is deemed to be on a target, not clutter, and is presented to the operator). Approaches such as those described in [5-10], occupy the middle ground. In these 'hybrid' methods, local clutter statistics and detection characteristics (such as signal-to-noise ratio, peak shape and peak curvature) are generated by the detector and used by the tracker.

The high spectral and spatial (angular) resolution of optical sensors, along with their rapid scan rates, support the application of TBD or hybrid approaches; however, hybrid approaches are more appropriate when computing resources are restricted and when target Signal to Noise Ratio (SNR) is high. When operating in the visible spectrum, peaks are typically identified using statistical change detectors [11,12] to suppress clutter (on background features of no interest). Trackers based on algorithms such as Probabilistic Data Association (PDA) [5-8,13], Joint Probabilistic Data Association (JPDA) [9,14], or Multiple Hypothesis Tracking (MHT) [10,15,16], are then applied.

In this communication a statistical change detector, implemented as a zero-latency finite-memory filter, is used to identify anomalies in temporal pixel statistics. An F-distributed test statistic is computed for each pixel and used in a hypothesis test. The tracker uses a pairwise JPDA algorithm, which has been restricted to consider clusters (sub-problems) containing no more than two tracks. This design allows multiple closely-spaced targets to be tracked at a low computational cost, relative to conventional JPDA. It uses the confidence model described in [5] to automatically confirm and delete tracks. The track state is also augmented to include color and the color statistics of the background (clutter) are estimated. The techniques described are applied to a sequence of color images, they are however, when restricted to a single color, equally applicable to any digital sensor, such as radar or sonar.

2. DETECTOR

The change detector used here does not operate on a difference map (i.e. the intensity difference between consecutive frames) [11]; rather, it estimates the parameters of a General Linear Model (GLM) [17], which is used to describe the time dependence of each color in each pixel. This detector primarily uses temporal statistics, rather than spatial statistics. The characteristics of the linear model are selected to best suit the behavior of the image and the capabilities of the processing platform. A model of arbitrary order may be used, employing fitting functions of arbitrary form; however, a zero-order constant model was found to be optimal, for the video data collected during algorithm development. When this approach is adopted, and the appropriate data structures implemented (circular buffers), the mean (the model parameter) and the variance can be computed efficiently using sliding windows to evaluate moving averages. The computational effort is then independent of the length of the sliding (analysis) window, as only the image intensity values entering and exiting the window are processed. An expression for the F-distributed test statistic is given below. Appropriate expressions for the more general case can readily be derived; however, due to space constraints, only the simplest (and recommended) case is provided here.

Under the null (no-change) hypothesis, it is assumed that the image intensity I , for the k th color of the pixel in the i th row and the j th column of a sequence of images, is distributed as a Normal variable, with constant $\mu(i, j, k)$, which is independent of time. It is also assumed that the variance σ^2 , is constant with respect to time, and equal for all colors and pixels, over the entire image, i.e.

$$I(i,j,k) \sim \mathrm{N}\{\mu(i,j,k), \sigma^2\} \tag{1}$$

The intensities are assumed to be independently distributed, with respect to all indices. The temporal constancy of the distribution parameters is assumed only over the recent time history. Using the N most recent measurements at the t th frame, for a given pixel and color, the Maximum Likelihood Estimates (MLEs) – $\hat{\mu}(i,j,k)$ and $\hat{\sigma}^2$ – of the (unknown) true parameters – $\mu(i,j,k)$ and σ^2 – of the intensity distributions are computed using

$$\hat{\mu}_t(i,j,k) = \frac{1}{N}\sum_{t'=t-N+1}^{t} I_{t'}(i,j,k) \tag{2}$$

and

$$\hat{\sigma}_t(i,j,k)^2 = \frac{1}{N}\sum_{t'=t-N+1}^{t} I_{t'}(i,j,k)^2 - \left\{\frac{1}{N}\sum_{t'=t-N+1}^{t} I_{t'}(i,j,k)\right\}^2 . \tag{3}$$

Under the null hypothesis, the following statistics may then be formed, upon receiving the next frame:

$$\{I_{t+1}(i,j,k) - \hat{\mu}_t(i,j,k)\}^2 \Big/ \left\{\left(1+\tfrac{1}{N}\right)\sigma^2\right\} \sim \chi^2\{1\} \tag{4}$$

and

$$N\hat{\sigma}_t(i,j,k)^2 / \sigma^2 \sim \chi^2\{N-1\} . \tag{5}$$

They are however of limited practical use because σ^2 is unknown. Summing the (4) and (5) statistics above, over all K colors, for a given pixel, also results in chi-squared distributions with K and $K(N-1)$ degrees of freedom respectively, due to the reproductive property of chi-squared variables. Dividing the former sum by the latter sum cancels σ^2; furthermore, dividing each by their respective degrees of freedom yields the desired F-distributed test statistic:

$$Z_t = \frac{(N-1)\sum_{k=1}^{K}\{I_{t+1}(i,j,k) - \hat{\mu}_t(i,j,k)\}^2}{(N+1)\sum_{k=1}^{K}\hat{\sigma}_t(i,j,k)^2} \sim \mathrm{F}\{K, K(N-1)\} . \tag{6}$$

The Z_t test statistic is evaluated at every pixel forming a (monochrome) map and the null hypothesis is tested. A large Z_t value means that an uncharacteristically large change in pixel intensity, for one or more colors, has been observed. Use of a general linear model permits the 'normal' behavior of each pixel to be adaptively determined; while the use of a test statistic with a known distribution allows the false alarm rate to be known and constant, when the assumptions underlying the null hypothesis are indeed true.

In practice, a low-pass (moving average) spatial filter is used, to remove 'speckle' noise in the map (yielding B_t), which tends to 'smear' features. The statistic is depleted for isolated large values of Z_t and accumulated when many large values are in close proximity. This filter makes it easier to identify discrete (moving) objects in the image and reduces the probability of producing multiple peaks on a single target. A threshold is selected to give the desired false-alarm rate, using the distribution of $\mathrm{F}\{K, K(N-1)\}$ as a guide. All values above the threshold set the corresponding pixels in the change mask to true, forming contiguous 'blobs' in the change mask, due to each moving object. The maximum of each blob is identified (a peak) and every true pixel in the change mask is assigned to a peak. For each peak, the mean of each color in the underlying image is computed, using the pixels in the most recent frame, as selected by the change mask. The mean color values are appended to the measurement vector. The mean and variance of the image in the vicinity peak are also estimated and used to parameterize the local clutter model.

3. TRACKER

A pictorial representation of the JPDA event space, extended to include target visibility [5], for two tracks and two common peaks (i.e. gated by both tracks) is given in Figure 1. The 'event space' defines the physical set of real-world assumptions used within the track updater's probabilistic Bayesian model.

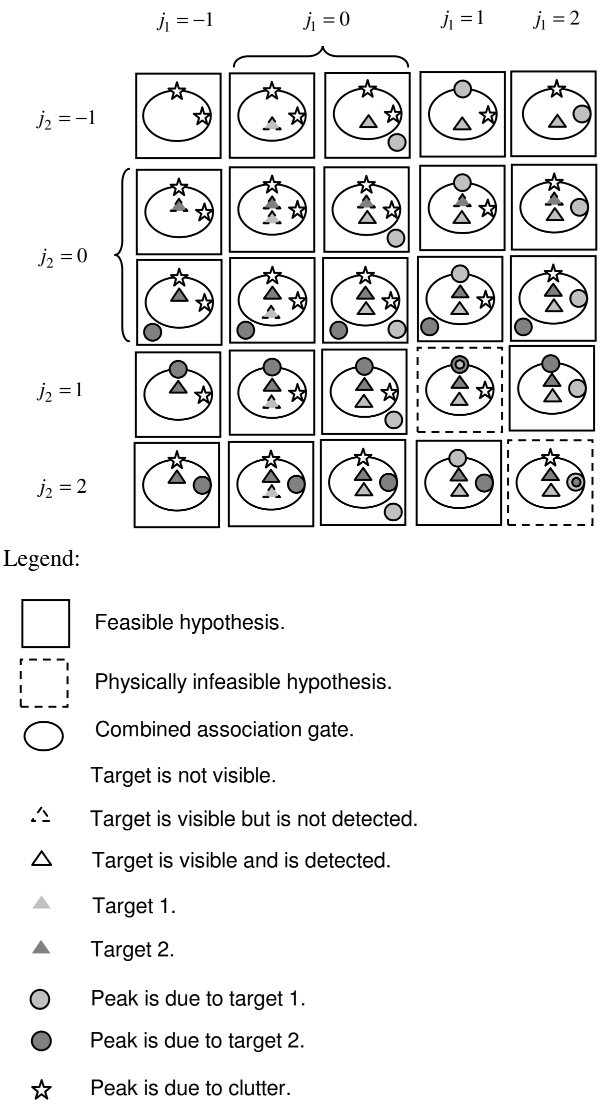

FIGURE 1. The JPDA event space.

Using the above diagram as a guide, the event probabilities, for two tracks, 'competing' for any number of common peaks, are generated using the following (approximate) expressions:

$$\beta_1'(j_1) = \beta_1(j_1) \sum_{j_2=-1}^{N_2} \beta_2(j_2), \; j_1 = -1 \ldots 0\,; \tag{7}$$

$$\beta_1'(j_1) = \beta_1(j_1) \sum_{j_2=-1}^{N_2} \left\{ \beta_2(j_2) - \delta_{j_1 j_2} \beta_2(j_2) \right\}, \; j_1 = 1 \ldots N_1\,; \tag{8}$$

$$\beta_1''(j_1) = \beta_1'(j_1) \Big/ \sum_{j_1=-1}^{N_1} \beta_1'(j_1)\,. \tag{9}$$

In the expressions above, subscripts 1 and 2 denote the 1st and 2nd tracks respectively, j is the event index, N is the number of peaks associated by a given track, β is the PDA event probability, β' and β'' are the un-normalized and normalized pairwise JPDA event probabilities (respectively), and δ is the kronecker delta function. The event probabilities of the 2nd track are computed by exchanging the subscripts. The physically infeasible events, corresponding to a single peak being due to different targets, are handled using the 'exchange interaction' term, $\delta_{j_1 j_2} \beta_1(j_1) \beta_2(j_2)$, in (8).

This (pairwise) method was adopted because it is fast to implement and execute, relative to (complete) JPDA. Restricting the algorithm to (at most) two tracks and targets (i.e. the objects being tracked) greatly reduces the complexity of the JPDA implementation, yet it is sufficient to resolve most commonly arising cases of association ambiguity, even if more than two tracks and two targets are involved. The more general case, for an arbitrary number of tracks and targets is handled as follows: The pair of tracks (if any) with the greatest association ambiguity is identified, by examining the calculated β values, and updated first using (7)-(9). Of the remaining tracks, the track pair (if any) with the next greatest association ambiguity is subsequently identified and processed, and so on, until all tracks have been updated. The PDA update is used when no association ambiguity exists. In terms of average tracking errors, this method is equivalent to PDA and JPDA for well-separated tracks; and superior to PDA and similar to JPDA for two (or possibly more) closely spaced tracks. To further reduce the processing load, only confirmed tracks are processed using the pairwise JPDA update. Tentative tracks are updated using the PDA update, using peaks that were not used to update the confirmed tracks. New tentative tracks are initiated on all peaks that were not used to update the confirmed and tentative tracks.

The idea of state augmentation for the purpose of target feature tracking has been taken from [5] and [6], where SNR and curvature of radar peaks (respectively) are used as discriminatory target features. For video data, when the target color is different to other nearby targets and the background, this technique has the potential to reduce: the incidence of track divergence, false track rates, and track confirmation times.

4. RESULTS

In a field experiment, approximately 3000 frames of size 240x320, with 8bit RGB color, were acquired at a rate of 7.5Hz using a digital video camera, then post-processed using the detector and tracker described in the preceding sections. The data set contains the following: multiple maneuvering and closely-spaced moving targets (vehicles), targets occluded by vegetation, a small amount of camera shake, tree tops perturbed by a gentle breeze and moving clouds causing illumination to vary. A single example from this set is presented below in Figures 2 – 4. The selected footage contains 3 vehicles simultaneously approaching, and converging upon, a 'T'-intersection (1 yields and 2 pass). As target 3 passes target 1, only one peak is produced by the detector. Figure 4 shows that the targets are successfully tracked through the intersection using the pairwise JPDA tracker. The upper left insert shows the moment of maximum association ambiguity, near the intersection, for the pairwise JPDA tracks. When the pairwise JPDA correction logic is disabled, the tracker reduces to PDA; in this case (upper right insert), track 3 is 'seduced' by the peaks supporting track 2; track 3 remains at the intersection while its target continues down the road, i.e. the track diverges.

FIGURE 2. Input image frame, I_t.

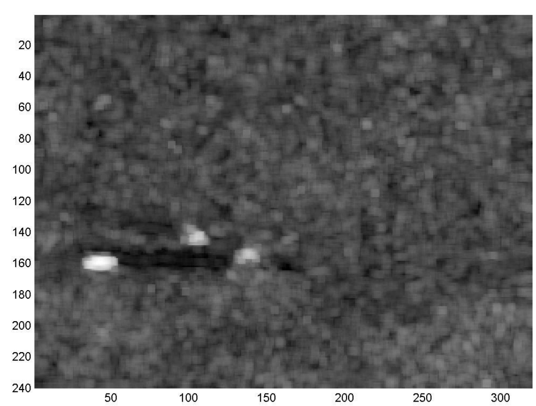

FIGURE 3. Detector output, $10\log_{10}(B_t)$.

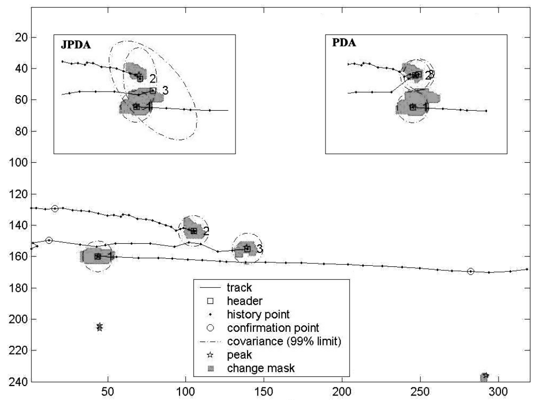

FIGURE 4. Tracker output.

5. DISCUSSION

The temporal detector used here does not yield peaks on object departure and object arrival, for low frame rates, as is the case when a difference map (between the current and previous images) is used. However, the temporal memory and the adaptive characteristics of this detector mean that the detection probability is reduced for a given target when it closely follows another target on the same trajectory, because the estimated variance is temporarily increased in the wake of the leading target (visible as dark smears behind the targets in Figure 3). The reliance of this detector on temporal statistics, rather than spatial statistics, results in higher spatial resolution, at the expense of a higher rate of false-peak production. Its tendency to produce false peaks on moving background objects, and during sudden changes in illumination and camera parameters (pan, tilt, zoom, translation and aperture) is, however, effectively handled by the tracker, which suppresses tracks on uncorrelated peak sequences, keeping the overall system false-alarm rate low. The use of more computationally intensive correlation-based [18] and gradient-based [19] optical flow techniques is likely to yield fewer false peaks when camera parameters are changed. However, for a given processor, increased computational load means a lower real-time frame rate, which makes tracking closely-spaced moving targets in clutter a more difficult task, especially when they are maneuvering.

6. CONCLUSION

A change detector and a target tracker, suitable for processing moving objects in color image sequences have been presented. The detector identifies anomalies in the temporal color statistics of each pixel. The tracker employs a pairwise JPDA algorithm, which minimizes track divergence for closely spaced tracks and the problem of multiple tracks on the same target. The tracking filter is extended to consider target visibility – for automatic track initiation and termination – and target color features, for improved track continuity.

7. REFERENCES

[1] Schleher, D. Curtis (Ed.), *Automatic Detection and Radar Data Processing*, Artec House, Dedham, MA, 1980.

[2] G.W. Pulford, "Taxonomy of multiple target tracking methods", *IEE Proc. Radar, Sonar and Navig.*, vol. 52, no. 5, pp. 291-304, Oct. 2005.

[3] Wei Zhang, Mingyu Cong and Liping Wang, "Algorithms for optical weak small targets detection and tracking: review", in *Proc. 2003 Int. Conf. on Neural Networks and Signal Process.*, vol. 1, pp. 643-647, Dec. 2003.

[4] P. Wei, J. Zeidler and W. Ku, "Analysis of multiframe target detection using pixel statistics", *IEEE Trans. Aerospace and Electronic Systems*, vol. 31, no. 1, pp. 238-247, Jan. 1995.

[5] S.B. Colegrove and S.J. Davey, "PDAF with multiple clutter regions and target models", *IEEE Trans. Aerospace and Electronic Systems*, vol. 39, no. 1, pp. 110-124, Jan. 2003.

[6] S.B. Colegrove, B. Cheung, "A peak detector that picks more than peaks", in *Proc. RADAR 2002,* pp. 167-171, Oct. 2002.

[7] T. Kirubarajan and Y. Bar-Shalom, "Low observable target motion analysis using amplitude information", *IEEE Trans. Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1367-1384, Oct. 1996.

[8] D. Lerro and Y. Bar-Shalom, "Interacting multiple model tracking with target amplitude feature", *IEEE Trans. Aerospace and Electronic Systems*, vol. 29, no. 2, pp. 494-509, Apr. 1993.

[9] H.M. Shertukde and Y. Bar-Shalom, "Tracking of crossing targets with imaging sensors", *IEEE Trans. Aerospace and Electronic Systems*, vol. 27, no. 4, pp. 582-592, Jul. 1991.

[10] S.S. Blackman, R.J. Dempster and T.J. Broida, "Multiple hypothesis track confirmation for infrared surveillance systems", *IEEE Trans. Aerospace and Electronic Systems*, vol. 29, no. 3, pp. 810-824, Jul. 1993.

[11] R.J. Radke, S. Andra, O. Al-Kofahi and B. Roysam, "Image change detection algorithms: a systematic survey", *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294-307, Mar. 2005.

[12] T.J. Patterson, D.M. Chabries and R.W Christiansen, "Detection algorithms for image sequence analysis", *IEEE Trans. Acoustics, Speech, and Signal Process*, vol. 37, no. 9, pp. 1454-1458, Sep. 1989.

[13] Y. Bar-Shalom, T. Kirubarajan and X. Lin, "Probabilistic data association techniques for target tracking with applications to sonar, radar and EO sensors", *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, pt. 2, pp. 37-56, Aug. 2005.

[14] T. Fortmann, Y. Bar-Shalom and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association", *IEEE J. Oceanic Engineering*, vol. 8, no. 3, pp. 173-184, Jul. 1983.

[15] S.D. Blostein and H.S. Richardson, "A sequential detection approach to target tracking", *IEEE Trans. Aerospace and Electronic Systems*, vol. 30, no. 1, pp. 197-212, Jan. 1994.

[16] F. Fukushima, M. Ito, S. Tsujimichi, and Y. Kosuge, "Track-initiation with the measurements of optical sensors by using MHT techniques", in *Proc. SICE 38th Annual Conference*, pp. 1123-1128, Jul. 1999.

[17] L.J. Bain and M. Engelhardt, *Introduction to Probability and Mathematical Statistics*, 2nd Ed., California, Duxbury Press, 1992.

[18] Teahyung Lee and D. Anderson, "Performance analysis of a correlation-based optical flow algorithm under noisy environments", in *IEEE Int. Symp. on Circuits and Systems, 2006. (ISCAS 2006)*, pp. 21-24, May 2006.

[19] M. Mesbah, "Gradient-based optical flow: a critical review", in *Int. Symp. on Signal Processing and Its Applications, (ISSPA '99)*, vol. 1, pp. 467-470, Aug. 1999.