18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

Fusion of Trackers on Thermal Image Sequences

Sebastian Thomé, Eckart Michaelsen, Norbert Scherer-Negenborn, Klaus Jäger, Leo Doktorski
Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB,
Gutleuthausstrasse 1, 76275 Ettlingen, Germany.
Sebastian.Thome@iosb.fraunhofer.de

Abstract – A set of trackers operating on thermal video data is investigated. The contribution focuses on marine scenarios with ships that may perform maneuvers such as turning or occluding each other. The tracker methods are not original, and they are not improved here, only their parameters were adjusted in order to fit the scenario. This contribution focuses on the fusion of the trackers. Their failures occur at different times on the same video. Thus, a committee solution performing a weighted mean is calculated and outlier-trackers are removed appears appropriate. The outliers are then newly locked on to the majority estimate. Thus, later on, the fusion may still benefit from their contribution. Experiments are performed on a thermal video taken at a Singapore waterfront displaying boats performing turning maneuvers and occluding each other.

Keywords: Thermal videos, tracking, fusion, data association.

1 Introduction

Various tracker algorithms have been proposed for tracking extended objects such as ships or boats on thermal image sequences. In case of correct tracking they all match. More interesting is the case of failure, i.e. loss of track. Here the different tracker algorithms sometimes fail at different time instances (on the same image sequence) and for different reasons.

Thus, it is an obvious idea to have a set of different tracker algorithms running in parallel on the same data and searching for criteria how to fuse their output, in order to improve the overall performance and minimize the failure probability.

The contribution at hand concentrates on thermal imagery. On such data occlusion of vehicles (land, water or air-based) often appears as merging of bright regions. As a result it is not immediately evident which target is occluded, and which target occludes the other.

1.1 Related Work

In [1] a framework was presented for fusion of point-based trackers and region-based trackers. The framework was demonstrated on a thermal image sequence containing boats partially occluding each other. As early as [2] work on tracking segments in thermal video data is discussed. The example sequence contains land-vehicles partially occluding each other. Tracking is done rather heuristically but the modelling of events – such as approaching, merging, occluding, splitting, or saliently parallel movement are handled explicitly using artificial intelligence methods. Among the multiple model trackers the interacting multiple model (IMM) provides an example of fusion by weighted sum [3,4,5]. These approaches run several estimators in parallel in order to cope with non-stationary target dynamics – e.g. changing steady courses to curve maneuvers. For the contribution at hand the filter (with its assumptions on the model dynamics) remains the same for all models in the pool, while the detection methods are divers.

The trackers used in this contribution have also been used as examples in an attempt to define a quantitative characteristic for the performance of trackers in general [6].

2 Method

A tracker consists of two steps: Prediction and detection. State-of-the-art are Kalman filters so as to properly combine the prediction (i.e. prior) with the detection (i.e. likelihood) in order to achieve a maximum a posteriori estimation of the target parameters (position, speed, etc.). The maximum a posteriori is estimated at each time step (i.e. incoming frame). This is commonly called *filtering*. There are linear methods and non-linear methods for such filtering. Most widely used is the alpha-beta-filter, see e.g. [7] Section 6.5.3.

Often the mass of the prior is strongly concentrated on a small image region – the *focus-of-attention*. For simplicity many methods code it as a rectangular search window with uniform distribution inside. Thus the prior density is approximated by giving only two image coordinates: upper left and lower right corner.

Different tracker methods differ in what they detect from the image, and in what information they hold and update. The following subsections list a handful of such methods and explain their specific properties. This is just for recall; the trackers are not the original contribution of this work.

This study also omits the first detection of the target – the *lock-on*. The lock-on method is assumed to work properly for the remainder of the paper. For simplicity it is done manually.

2.1 The reticle tracker

Detection: A reticle system consists essentially of a lens, a rotating mask placed in the image plane and a detector. The mask is also called reticle and blocks certain portions of the scene. The detector generates a 1D-signal over time that encodes the coordinates of the object. Fig. 1 shows a schematic representation of a reticle system. Further information can be found in [8], [9], [10], [11] and [12].

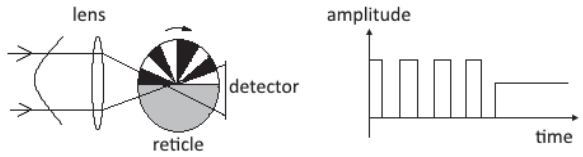

Figure 1. Left: Schematic representation of a reticle system. Right: Amplitude of the generated Signal.

Our tracker simulates a reticle system by masking the

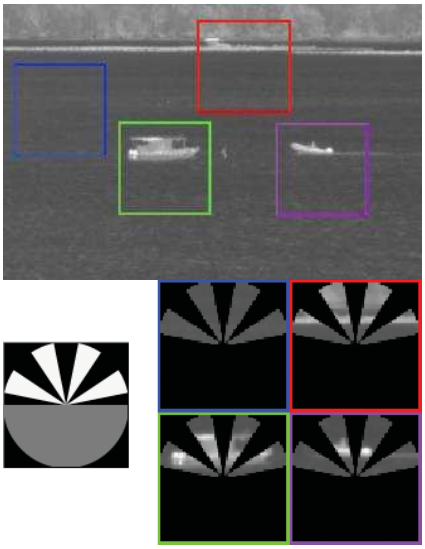

Figure 2. Top: thermal image with marked sub-images. Below: Reticle mask with 4 fans and masked sub-images.

image with the reticle and calculating the integral from the FOV. Fig. 2 shows a thermal image and a reticle mask.

Focus-of-attention: Circular and fixed-size.

Model: Position and speed of the target in 2D.

Filtering: Using an alpha beta filter.

Parameters: Number of fans, rotation speed of the reticle, scale of antialiasing to calculate the mask, alpha and beta value for the filter.

2.2 The hot-spot tracker

Detection: The hot-spot tracker is a simple tracker, which follows the hottest point in its field of view. In marine scenarios a ship with its engine or funnel is often the hottest object in the image. So in simple scenarios with only one ship this tracker often succeeds. But this tracker cannot differentiate two ships that overlap. The hot-spot method looks for the brightest pixel in its focus-of-attention – a simple max-operation.

Focus-of-attention: Rectangular and fixed-size.

Model: Position and speed of the target in 2D.

Filtering: Using an alpha beta filter.

Parameters: Alpha and beta value for the filter.

2.3 The centroid tracker

Detection: A centroid tracker calculates the center of mass of its search region and uses this center as the new position of the object. We use a binary centroid tracker which preprocesses its search region with a binary segmentation. Therefore a threshold is needed to segment the foreground from the background. We assume that the lock-on in the first image is done properly and this threshold can be automatically calculated using histograms. Fig. 3 shows a preprocessed thermal image.

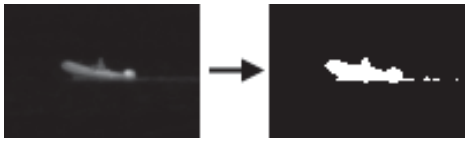

Figure 3. Left: thermal image before preprocessing. Right: thermal image after preprocessing.

Apart from position and speed of target, this tracker can also estimate the size of the object.

Focus-of-attention: Variable-size rectangular, if detection region is growing the focus-of-attention will also grow.

Model: Position, size and speed of the target in 2D.

Filtering: Using an alpha beta filter.

Parameters: maximal growing factor of object, maximal shrinking factor of object, threshold for segmentation, contrast type, alpha and beta value for the filter.

2.4 The fixed-correlation tracker

Detection: A correlation tracker tracks the object by matching the image with a predefined template. This matching is done with the normalized cross-correlation

$$corr = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}. \qquad (1)$$

Here x_i is a pixel's brightness in the image and y_i a pixel value in the template. Accordingly $\bar{x}$ is the mean of the image and $\bar{y}$ is the mean of the template. This correlation is calculated for every possible shift in the search region and the maximum value is the new object position. The template is rectangular.

If this value is under a certain threshold we assume that the object is lost. In this case the position will be predicted using the filter.

Fixed-correlation means that the template does not change over time, the template is cropped from the lock-on-image.

Focus-of-attention: Rectangular and fixed-size.

Model: Position, speed of the target in 2D.

Filtering: Using an alpha beta filter.

Parameter: threshold for correlation value, maximum count for dropping the object, alpha and beta value for the filter.

2.5 The average-correlation tracker

Detection: The only difference between the fixed- and the average-correlation tracker is that the template is variable over time. The new template is calculated by adding some information from the best match in the new image. This tracker can follow objects that change their appearance slowly over time, like turning ships.

Focus-of-attention: Rectangular and fixed-size.

Model: Template, position, size and speed of the target in 2D.

Filtering: Using an alpha beta filter.

Parameter: threshold for correlation value, maximum count for dropping the object, ratio of changing the template, alpha and beta value for the filter.

2.6 Fusion

The main idea here is in achieving robustness by forming a weighted mean according to

$$P_f(t) = \frac{\sum_{i \in I} w_i b_i(t) P_i(t)}{\sum_{i \in I} w_i b_i(t)} . \quad (2)$$

Here $P_i(t)$ is the position estimate of the i-th tracker in the tracker pool I at time (frame number) t. There are two weights: w_i is a time-independent measure of confidence in tracker i; and $b_i(t)$ is the assessment the i-th tracker provides for its current estimate at time t. It is known that for different reasons a tracker $j \in I$ may fail at some time t, i.e. wrong position estimates. This will lead to a strong deviation from the mean, which can be detected by use of a threshold th_o, and such outlier can be removed according to:

$$\left\| P_f(t) - P_j(t) \right\| > th_o \Rightarrow I \leftarrow I - \{j\} . \quad (3)$$

After such removal $P_f(t)$ is calculated again using eq. (2). If all $P_i(t)$ deviate more than th_o from the mean track loss has to be announced.

Such "committee of trackers" fusion requires several rationales: 1) How to decide that a tracker has lost track; in our case by outlier detection; 2) to what position a lost tracker is newly locked on; in our case to the mean taken over the inliers; 3) how to set the weights in eq. (2); in our case by estimates of experienced experts. The latter could

Figure 4. Section of example frame #56: The target (i.e. the boat on the right) is moving steadily to the left.

Figure 5. Section of example frame #165: Partial occlusion of the target by another larger vessel in the image center.

Figure 6. Section of example frame #469: The target (now left of the other vessel) is turning and thus changing its appearance.

be replaced by machine-learning methods, if sufficient representative data with ground-truth were available

3 Experiments

Section 3.1 presents a thermal video of a marine scene with several vessels taken in the thermal spectral domain. Section 3.2 explains how ground-truth was obtained for this video. The particular experiments performed on these data are outlined in Section 3.3. In particular, events are discussed such as when a particular member of the tracker-pool loses track, and why.

3.1 Real thermal video

Figs. 4, 5, and 6 show sections from frames of a video taken of the Singapore harbor area in April 2011. It is a long-wave infrared video (8-9µm) of dimension 640×480 pixel, taken by a cooled AIM QWIP-LW camera. It contains several vessels occluding each other now and then. It is recorded with 25 Hz and has 1200 frames.

3.2 Ground-truth

The movements of the vessels in such data are comparably slow and steady. Thus, it suffices to mark the target manually on e.g. every tenth image. This was done for this contribution. Of course less subjective methods such as automatic solutions would be preferable. But the data are challenging; ATR-method might fail e.g. because of occlusion and appearance change during maneuvers. The best solution would be precise Geo-references on the target (through GPS, INS and/or radio-localization) and a precisely calibrated camera setup.

3.3 Evaluation

In the current study we used the measure of confidence listed in Table. 1. The measure w_i of the fixed-correlation tracker is relatively high because this tracker is more reliable than the others. As assessment $b_i(t)$ for the two correlation-trackers we use their correlation value shown in eq. (1). For the other three trackers there is currently no dynamic assessment available so we selected them as constant 1. The threshold th_o was set to 60 pixels in order to give every tracker enough space.

For our experiments the trackers were manually locked on the maneuvering vehicle. Fig. 7 displays the Euclidean distance between tracking estimate and Ground-truth position as a function of time (frame #). It is obvious that a critical situation is caused by the occlusion. One by one the trackers of the pool fail. Only the *fixed-correlation* method can bridge this gap and perfectly locks on after predicting the correct position from dead reckoning the last observed speed before occlusion. The others are – one after another– removed from the pool, and newly locked on (which is not displayed in Figure 7).

The *fixed-correlation* method, however, has problems with the turning maneuver starting around #379. The appearance changes and it thus decides to neglect the detection and continue on the prediction still running to the left. At #550 it is detected as an outlier, removed from the pool, and newly locked on the mean estimate of the other methods. These are more robust with respect to slowly changing appearance.

A similar situation occurs again starting with #690. Now the target (still the small boat) is starting to occlude the other vessel. It is fairly typical for thermal data that in such occlusion event, it can hardly be distinguished what part is occluded, and what part is in the front. The objects have roughly the same temperature, so that the bright blobs just merge and no occluding boundary is detectable. Again, the methods fail one by one and only the *fixed-correlation* method remains on track. Again, it can establish re-detection after the occlusion situation is over at #880.

Starting from #930, when the boat's appearance changes again due to another turning maneuver, the *fixed-correlation* method loses track, while the others hold on. At #1070 it is again detected as an outlier, and the track is automatically locked on the mean estimate, which is again close to the ground truth.

The whole sequence completes one target's 'orbit' round the other vessel, and the fusion method is the only one considered here which can remain on target in this situation.

4 Discussion and Conclusion

The results clearly indicate that tracking may benefit well from such fusion approaches. There are reasons for the construction of different trackers. Some are good at finding a previously seen target again after comparably long occlusion; some are robust with respect to appearance changes, e.g. induced by target maneuvers; and so on.

Yet this fusion is not an easy task. Finding appropriate or optimal weights for formula (2) is not trivial. It may require expert knowledge or machine-learning on huge representative data sets.

Further investigations might focus on developing an assessment $b_i(t)$ for the reticle-, hot-spot- and centroid-tracker which have been defined as constant 1 in this study. A nonlinear function can provide better results. As a consequence the existence of a more efficient assessment than $b_i(t) = corr(t)$ for both correlation trackers might be of interest.

It can also be important to find out which trackers are directly related to a reliable result under what circumstances. Maybe rules can be formulated allowing automatic selection of preferences or weight choices.

Table 1: Confidence in tracker and assessment.

	reticle	hot-spot	centroid	fixed-correlation	avg-correlation
w_i	0.125	0.125	0.125	0.5	0.125
$b_i(t)$	1	1	1	corr(t)	corr(t)

Table 2: Frame numbers of track-loss: colors corresponding to Fig. 7.

Event	Frame #	reticle	hot-spot	centroid	fixed-correlation	avg-correlation
Steady moving	1 - 150	on track	on track	on track	on track	on track
Occlusion	151 - 295	loss	loss	loss	predicting	loss
Steady moving	296 - 378				on track	
Target turning	379 - 550				loss	

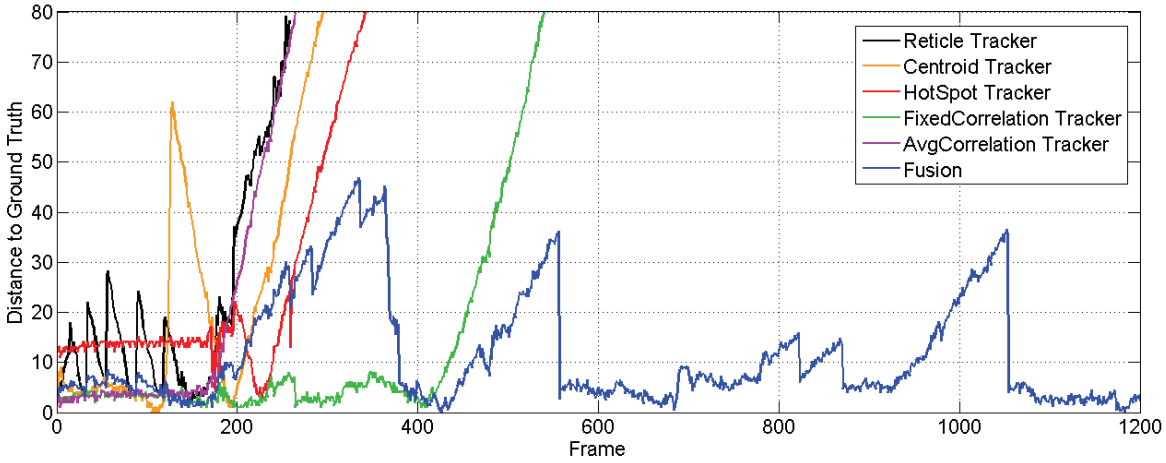

Figure 7. Behavior of the trackers in the pool: Euclidean distances to ground-truth as function of frame number

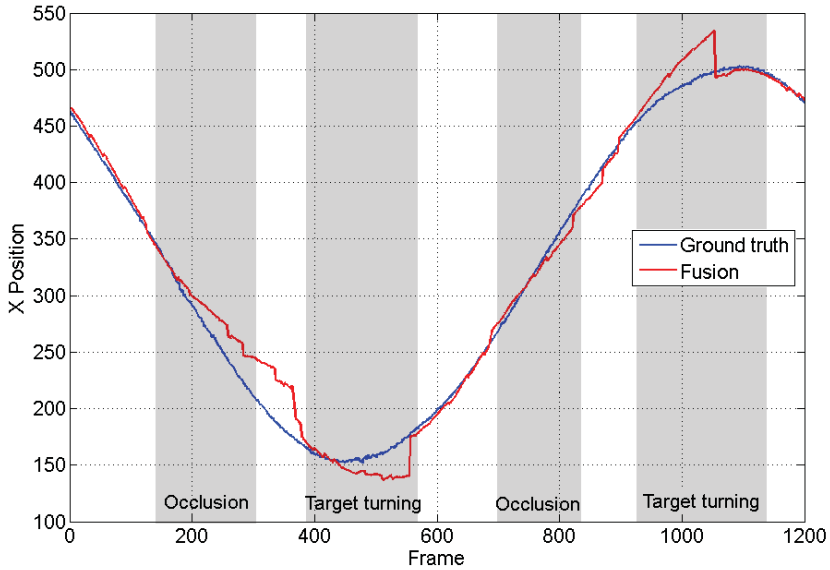

Figure 8. Deviation of the fused tracker: blue: Ground-truth, x-coordinate in Pixel as function of frame number; red: position estimate, again x-coordinate.

Acknowledgement

The real thermal video was taken as part of the Singapore-German cooperation KOSINUS contract number E/SR2I/AA214/9F192 in April 2011.

References

[1] Michael Teutsch et al., "Fusion of region and point-feature detections for measurement reconstruction in multi-target Kalman tracking," *14th International Conference on Information Fusion 2011*, Chicago, Illinois, USA, 2011, pp. 5-8.

[2] Ullrich Thönnessen, Dieter Ernst, Hermann Groß, "Entwicklung einer segmentbasierten Beschreibung von Ereignissen in Bildfolgen," *13th DAGM (German Conference on Pattern Recognition) 1991*, Munich, Informatik Fachberichte 290, Springer, Berlin, 1991, pp. 499-514.

[3] H. A. P. Blom, "An efficient filter for abrubtly changing systems," *23rd IEEE Conference on Decision and Control*, Las Vegas, 1984, pp. 656-658.

[4] E. Mazor, A. Averbuch, Y. Bar-Shalom, J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Trans. on Aero. And Elec. Sys.,* vol. 34, no. 1, 1998, pp. 103-123

[5] Ludmila Mihaylova, Emil Semerdjiev, „An interacting multiple model algorithm for stochastic system control," *Information & Security, vol. 2,* 1999.

[6] Leo Doktorski, Eckart Michaelsen, Endre Repasi, "A single performance characteristic for the evaluation of seeker tracking algorithms," *Pattern Recognition and Image Analysis*, vol. 24, no. 2, pp. 218-225, 2014.

[7] Yaakov Bar-Shalom, X. Rong Li, Thiagalingam Kirubarajan: *Estimation with Applications to Tracking and Navigation*, John Wiley & Sons, New York, 2001.

[8] R. Legault, *The Infrared Handbook*, Chap. 17: Reticle and Image Analyses, Ed. by W. L. Wolfe and G. J. Zissis, Office of Naval Research, Department of the Navy, Washington, 1989.

[9] K. Suzuki, "Analysis of rising□sun reticle," Optical Engineering, vol. 18, no. 3, pp. 350-351, 1979.

[10] J.S. Accetta, D.L. Shumaker: *The Infrared and Electro-Optical Systems Handbook, Vol. I – VIII*, SPIE Optical Engineering Press, 1993.

[11] Richard D. Hudson Jr.: *Infrared System Engineering*. John Wiley & Sons, New York, 1969.

[12] William L. Wolfe, Georg J. Zissis: *The infrared Handbook*, Office of Naval Research, Department of the Navy, Arlington, VA, 1978.