Infrared Target Tracking with AM-FM Consistency Checks

Nick A. Mould, Chuong T. Nguyen, and Joseph P. Havlicek
School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019 USA

Abstract

Challenging infrared data sequences such as the well-known AMCOM closure sequences are characterized by highly nonstationary, evolutionary target and clutter signatures, poor target-to-clutter ratios, and complex kinematics arising from both the target motion and the motion of the sensor platform itself. In such cases, track consistency checks can provide a valuable means for detecting an imminent track loss. In this paper, we consider a simple target model with a correlation-based detection process and a straightforward SIR particle filter track processor. We show that the performance of the track processor can be dramatically improved by incorporating modulation domain consistency checks to identify failure in the correlation-based detection process. This strategy results in a robust dual-domain tracker that, despite the simplicity of its state model, delivers superior tracking performance against the very difficult AMCOM sequences.

1. Introduction

We consider the difficult problem of tracking extended targets in midwave (3 μm - 5 μm) and longwave (8 μm - 12 μm) infrared video sequences. We focus our attention on the well-known *AMCOM* closure sequences [4,9,16,17]. This data set consists of ~ 40 sequences where an airborne maneuvering sensor closes on a variety of moving and stationary terrestrial targets. In general, these data represent challenging tracking problems because the target signatures undergo significant magnification and pose variations. In addition, many of the sequences are characterized by strong, highly structured clutter and/or low target-to-clutter ratios.

A standard approach to the infrared tracking problem involves detection based on template correlation followed by an optimal Kalman estimator. However, in cases where the target state space model is nonlinear or the measurement and/or process noises are non-Gaussian, the Kalman filter is not optimal. Particle filtering [1–3, 15] has emerged recently as an empirical tracking approach that often performs better than the Kalman filter against such cases. Although it is possible within the particle filtering framework to combine the detection and tracking processes [2, 3], here we will adhere to the more traditional model where detection and tracking are considered separately.

We assume a midwave or longwave imaging sensor that acquires video frames f_k, $k \in \mathbb{N}$. We refer to the raw pixel values $f_k(x_1, x_2)$ as the *pixel domain* representation of the frame. Pixel domain detection is accomplished by template matching (*e.g.*, normalized correlation) [10, 14], where a target detection is declared at the peak of the normalized correlation function. The template is initialized by manually designating a window about the target in the first frame f_0. This approach affords generality in the sense that it could correspond to a signature match in a system employing a library of stored signatures or alternatively to an assisted targeting system with a human in the loop who manually designates a previously unknown target type.

Subsequent to the initial frame, the track processor is expected to run autonomously without further human intervention. Provided that the target signature does not exhibit significant temporal evolution, correlation detection and tracking tend to be effective since the initial template provides a reasonable characterization of the target throughout the video sequence. In scenarios such as the AMCOM sequences, however, the target signatures exhibit *significant* nonstationary evolution. Consequently, after only a few frames the template becomes stale in the sense that it no longer provides an accurate characterization of the observed target. When this occurs, the target is typically lost and the tracker often locks onto structured features of the clutter, an effect that is exacerbated by the apparent motion of the background in the image plane that arises from the sensor platform kinematics.

In the literature, this has been referred to as the *template update problem* [10]. As the target signature evolves, the template must be updated or refreshed in order to preserve the quality of the detection process and maintain track lock. A naïve approach is to refresh the template every $L \geq 1$ frames based on the target signature observed in the last tracked frame. However, for highly evolutionary target signatures one must choose L small. Inevitably, the template overadapts and becomes matched to the background instead of the target. A more robust approach is to update the template when the observation fails to agree with the predicted track centroid. This approach still fails against highly evolutionary signatures, however, since the quality of the prediction is degraded rapidly as the template becomes stale.

In this paper, we introduce powerful new AM-FM consistency checks to identify when the detection process has been compromised by target signature evolution. These new checks, which are based on a modulation domain correlation tracker, are integrated with a pixel domain particle filter to obtain a robust new dual-domain track processor where stale templates are detected and refreshed on the fly. As illustrated by the examples given in Section 4, this integrated approach has proved remarkably effective against the AMCOM sequences. In Section 2 we briefly describe computation of the modulation domain image model, while details of the dual-domain track processor are given in Section 3.

This work was supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under grant W911NF-04-1-0221.

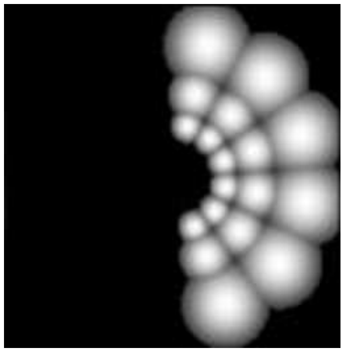

Figure 1. Frequency response of 18-channel Gabor filterbank for infrared target tracking.

2. Modulation Domain Frame Model

Our key hypothesis is that degradation of the pixel domain target track can be detected by simultaneously observing the target and background in a second domain where the target signature evolution, and hence failure of the track processor, is manifest differently. AM-FM models [5, 8, 13] provide a modulation domain image representation in terms of localized amplitude and frequency modulations. Recently, we demonstrated that such models are effective for representing infrared targets and backgrounds and can substantially enhance target-clutter class separability [7, 11, 12].

Since the AM and FM functions of a real-valued image component are ambiguous, we regularize the demodulation problem by constructing a complex multicomponent model t_k for each infrared video frame f_k according to

$$t_k(x_1, x_2) = f_k(x_1, x_2) + j\mathcal{H}[f_k(x_1, x_2)] \quad (1)$$

$$\approx \sum_{m=1}^{M} a_m(x_1, x_2) \exp[j\varphi_m(x_1, x_2)], \quad (2)$$

where $\mathcal{H}[\cdot]$ is the partial Hilbert transform [6]. It is unnecessary for (2) to provide a complete representation of the frame f_k. Rather, what is important is that the M AM-FM components capture the structurally important features of the targets and backgrounds. Through extensive experimentation, we determined empirically that representation in terms of the $M = 18$ frequency selective Gabor filterbank channels shown in Fig. 1 is sufficient. Filterbanks of this type were described in [5, 8].

After applying the filterbank to the complex image t_k, we estimate the modulating functions a_m and $\nabla\varphi_m$ in (2) from the filterbank channel responses y_m according to [5]

$$\nabla\varphi_m(x_1, x_2) = \mathrm{Re}\left[\frac{\nabla y_m(x_1, x_2)}{j y_m(x_1, x_2)}\right] \quad (3)$$

and $a_m(x_1, x_2) = |y_m(x_1, x_2)|$. Intuitively, the AM functions a_m capture the local envelope, while the FM functions $\nabla\varphi_m$ capture local orientation and pattern spacing. The AM-FM model for f_k consists of the 54 images a_m, r_m, and θ_m, $1 \le m \le 18$, where $r_m = |\nabla\varphi_m|$ and $\theta_m = \arg \nabla\varphi_m$ are the polar representation of the vector field $\nabla\varphi_m$.

3. Dual-Domain Track Processor with AM-FM Consistency Checks

The pixel domain target centroid and velocity are modeled with a four-element state vector $\mathbf{x}_k = [x_{1,k}\ \dot{x}_{1,k}\ x_{2,k}\ \dot{x}_{2,k}]^T = [\mathbf{x}_{1,k}^T\ \ \mathbf{x}_{2,k}^T]^T$, where $\mathbf{x}_{1,k} = [x_{1,k}\ \ \dot{x}_{1,k}]^T$ and $\mathbf{x}_{2,k} = [x_{2,k}\ \dot{x}_{2,k}]^T$. We employ a constant velocity model, which is justified by the fact that the sensor platform closes on the target in the AMCOM sequences. The state transition equation is given by

$$\begin{bmatrix} \mathbf{x}_{1,k+1} \\ \mathbf{x}_{2,k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,k} \\ \mathbf{x}_{2,k} \end{bmatrix} + \mathbf{v}_k, \quad (4)$$

where

$$\mathbf{F}_1 = \mathbf{F}_2 = \begin{bmatrix} 1 & \Delta \\ 0 & 1 \end{bmatrix}, \quad (5)$$

Δ is the interframe time, $\mathbf{v}_k = [0\ v_{1,k}\ 0\ v_{2,k}]^T$, and $v_{1,k}$ and $v_{2,k}$ are zero-mean white Gaussian noises uncorrelated with one another. The pixel domain observation model is given by

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k, \quad (6)$$

where

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad (7)$$

$\mathbf{n}_k = [n_{1,k}\ n_{2,k}]^T$, and $n_{1,k}$ and $n_{2,k}$ are zero-mean white noises uncorrelated with one another and with $\mathbf{v}_k$.

We process the pixel domain observations with a standard SIR particle filter [1, 3, 17], where the state transition prior follows from (4), the likelihood of the observation $p(\mathbf{z}_k|x_k^i)$ follows from (6) with $n_{1,k}$ and $n_{2,k}$ assumed to be Rayleigh distributed, and resampling is performed at each time step. The Rayleigh assumption on the measurement noise was justified by empirical analysis of several AMCOM sequences against Monte Carlo simulations of the state transition equation. The pixel domain tracked centroid is given by the expectation of $[x_{1,k}\ x_{2,k}]^T$ with respect to the current particle population and weights, where the weights are proportional to the likelihood.

Performance of the pixel domain SIR track filter depends critically on the quality of the observations $\mathbf{z}_k$, which deteriorates rapidly if the template is permitted to become stale. In order to detect when a template refresh is needed, we run a modulation domain normalized correlation tracker in parallel with the pixel domain track filter. Each modulation domain pixel is 54-element vector comprising scalar entries from the 54 images a_m, r_m, and θ_m. Thus, one may visualize the modulation domain image representation as a cube constructed by stacking the 54 images a_m, r_m, and θ_m.

The modulation domain template has the same horizontal and vertical extent as the pixel domain template. However, like the modulation domain image representation, it consists of pixels that are 54-element vectors. The modulation domain template is initialized in the first frame by drawing 54-element vectors from the images a_m, r_m, and θ_m at the same pixel sites that are used to initialize the pixel domain template. Normalized correlation is performed in the modulation domain "cube," and the modulation domain track centroid is declared at the global maximum of the modulation domain correlation function.

A stale template condition is readily detected by divergence of the pixel domain and modulation domain track centroids. Due to nonstationary evolution of the target signature in the AMCOM sequences, there are instances where the pixel domain tracker fails first, where the modulation domain tracker fails first, and where both trackers fail simultaneously. However, it is extraordinarily rare that the failure modes in the two domains are similar. Thus, an impending track loss in one or both domains is almost always accompanied by a divergence of the two centroids.

When divergence is detected, we backtrack to the last frame where the centroids agreed and refresh both the pixel domain and modulation domain templates. We also consider a variety of hypotheses regarding the size and aspect ratio of the template. A fixed list of increases and decreases are applied to both the horizontal and vertical template dimensions. Each hypothesis is evaluated based on the first four moments of the template pixel values. The hypothesis yielding minimum L_2 distance to the moment vector of the observed target signature in the last frame where the pixel domain and modulation domain centroids agreed is accepted.

4. Results and Discussion

In this Section, we present results for two longwave infrared AMCOM closure sequences. As a baseline for comparison, we implemented a standalone pixel domain normalized correlation tracker with a fixed template, with template update in every frame, and with a variety of fixed interval template update strategies. None of these was capable of delivering satisfactory performance against the AMCOM sequences. For illustrative purposes, we show here the results for the fixed template strategy.

One of the most difficult AMCOM sequences is *rng19_13*, which depicts a convoy of three vehicles rounding a corner. In addition to significant magnification and aspect variations that result from closure of the sensor platform, the target signatures in this sequence undergo profound evolution due to the kinematics of rounding the corner. A pair of frames, one early and one late, are shown in Fig. 2, where the tracked centroid is shown as a black cross. Results for the dual-domain tracker with AM-FM consistency checks are given in Fig. 2(a) and (b). In both cases, the SIR filter track centroid is located on the lead target (as it is throughout the entire sequence). Results for the standalone pixel domain tracker are given in Fig. 2(b) and (d). In this case, the lead vehicle is lost as soon as it turns the corner and the tracker variously locks onto background clutter and following vehicles as they come near. Errors in the detection process relative to ground truth for both the dual-domain and standalone pixel domain trackers are given for the entire *rng19_13* sequence in Fig. 3.

Results for a second, more typical AMCOM run (*rng16_07*) are given in Fig. 4. The target in this case is a slow moving truck and there are significant magnification variations due to closure of the sensor platform. SIR track centroids for the dual-domain track processor are given in the upper row (Fig. 4(a)-(d)) for frames 250, 300, 350, 360, while the track centroids for the standalone pixel domain correlation tracker are given in Fig. 4(e)-(h) for the same frames. As is typical, the pixel domain tracker totally fails to maintain track lock in this case. By contrast, the SIR filter centroid from the dual-domain track processor remains on the main target signature throughout the entire sequence.

In this brief paper, we have demonstrated that AM-FM consistency checks can dramatically improve the performance of elementary track processors against challenging, real-world infrared data sequences. Several important questions remain open, including that of incorporating the modulation domain observations directly into the measurement and state equations.

References

[1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online non-linear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, 50(2):174–188, Feb. 2002.

[2] M. Bruno. Bayesian methods for multiaspect target tracking in image sequences. *IEEE Trans. Signal Process.*, 52(7):1848–1861, Jul. 2004.

[3] M. G. S. Bruno. Sequential importance sampling filtering for target tracking in image sequences. *IEEE Signal Process. Lett.*, 10(8):246–249, 2003.

[4] A. Dawoud, M. Alam, A. Bai, and C. Loo. Decision fusion algorithm for target tracking in infrared imagery. *Opt. Eng.*, 44(2), 2005.

[5] J. P. Havlicek, D. S. Harding, and A. C. Bovik. Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models. *IEEE Trans. Image Proc.*, 9(2):227–242, Feb. 2000.

[6] J. P. Havlicek, J. W. Havlicek, and A. C. Bovik. The analytic image. In *Proc. IEEE Int'l. Conf. Image Proc.*, Santa Barbara, CA, Oct. 26-29, 1997.

[7] J. P. Havlicek, C. T. Nguyen, and M. Yeary. Modulation domain infrared target models. In W. R. Watkins and D. Clement, editors, *Targets and Backgrounds XII: Characterization and Representation*, volume 6239 of *Proc. SPIE*, pages 62390D–1 – 62390D–11, 2006.

[8] J. P. Havlicek, P. C. Tay, and A. C. Bovik. AM-FM image models: Fundamental techniques and emerging trends. In A. C. Bovik, editor, *Handbook of Image and Video Processing*, pages 377–395. Elsevier Academic Press, Burlington, MA, 2nd edition, 2005.

[9] J. Khan and M. Alam. Efficient target detection in cluttered flir imagery. In D. Casasent and T.-H. Chao, editors, *Optical Pattern Recogniction XVI*, volume 5816 of *Proc. SPIE*, pages 39–53, 2005.

[10] I. Matthews, T. Ishikawa, and S. Baker. The template update problem. *IEEE Trans. Pattern Anal., Machine Intel.*, 26(6):810–815, Jun. 2004.

[11] C. T. Nguyen and J. P. Havlicek. Modulation domain features for discriminating infrared targets and backgrounds. In *Proc. IEEE Int'l. Conf. Image Proc.*, pages 3245–3248, Atlanta, GA, Oct. 8-11, 2006.

[12] C. T. Nguyen, J. P. Havlicek, and M. Yeary. Modulation domain template tracking. In *Proc. IEEE Int'l. Conf. Comput. Vision, Pattern Recog.*, page 8 pp., Minneapolis, MN, Jun. 17-22, 2007.

[13] M. Pattichis and A. Bovik. Analyzing image structure by multidimensional frequency modulation. *IEEE Trans. Pattern Anal., Machine Intel.*, 29(5):753–766, May 2007.

[14] G. Stamou, M. Krinidis, E. Loutas, N. Nikolaidis, and I. Pitas. 2D and 3D motion tracking in digital video. In A. Bovik, editor, *Handbook of Image and Video Processing*, pages 491–517. Elsevier Academic Press, Burlington, MA, 2 edition, 2005.

[15] F. van der Heijden. Consistency checks for particle filters. *IEEE Trans. Pattern Anal., Machine Intel.*, 28(1):140–145, Jan. 2006.

[16] S. Yi and L. Zhang. A novel multiple tracking system for UAV platforms. In D. Henry, editor, *ISR Systems and Applications III*, volume 6209 of *Proc. SPIE*, 2006.

[17] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Comput. Surv.*, 38(4):1–45, Dec. 2006.

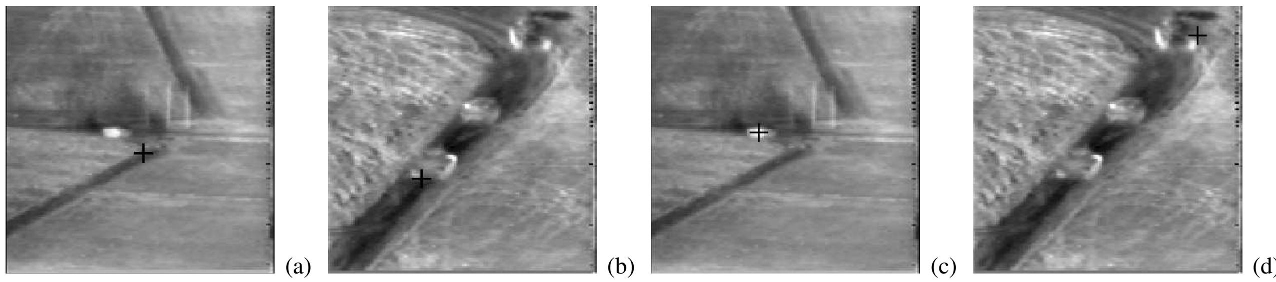

Figure 2. Tracking results for AMCOM run *rng19_13*. The problem is to track the lead vehicle around the turn. (a),(b) Dual-domain tracker at frames 150 and 250. (c),(d) Standalone pixel domain correlation tracker at frames 150 and 250.

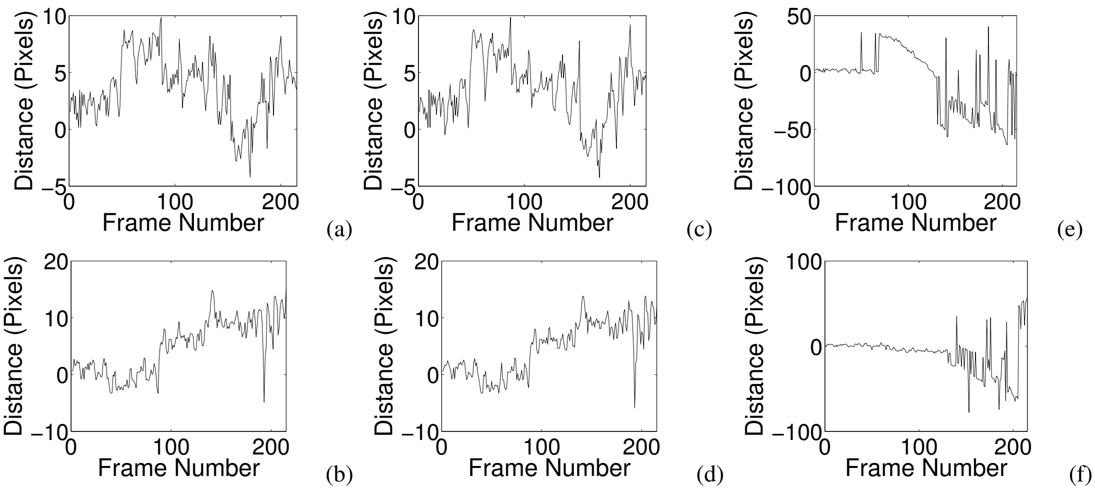

Figure 3. Horizontal and vertical detection errors (pixels) for AMCOM run *rng19_13*. Dual-domain tracker: (a),(b) x_1 and x_2 pixel domain correlation errors. (c),(d) x_1 and x_2 modulation domain correlation errors. (e),(f) Standalone pixel domain correlation tracker errors.

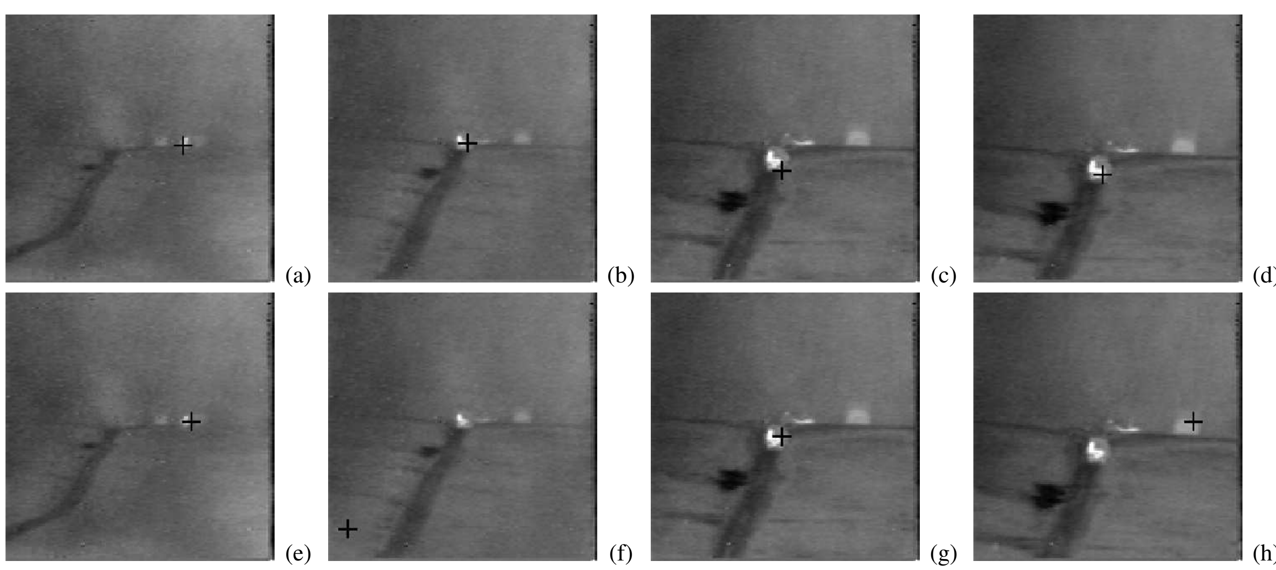

Figure 4. Tracking results for AMCOM run *rng16_07*, frames 250, 300, 350, and 360. (a)-(d) Dual-domain tracker. (e)-(h) Standalone pixel domain correlation tracker.