MULTICHANNEL DUAL DOMAIN INFRARED TARGET TRACKING FOR HIGHLY EVOLUTIONARY TARGET SIGNATURES

Colin M. Johnston, Nicholas A. Mould, and Joseph P. Havlicek

School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK, USA

ABSTRACT

We introduce a new SIR particle filter that performs tracking in a joint feature space where pixel domain data are fused with measurements obtained from an 18-channel modulation domain image model. This dual domain processor is capable of maintaining track lock and delivering a high probability of kill against targets in the AMCOM infrared closure sequences, which are challenging because they are characterized by high closure rate dynamics, poor SNR, and maneuvering targets that typically exhibit severe signature evolution and profound magnification changes. In the setup considered here, the track processor receives an initial target designation, which could be obtained from an external detector or from a human in the loop. After the first frame, the track processor must run autonomously without further *a priori* information. Compared to traditional pixel domain trackers, our results demonstrate that this new dual domain approach provides inherently improved tracking accuracy and facilitates powerful new consistency checks capable of detecting when a template update is needed due to nonstationary target signature evolution.

Index Terms— AMCOM, AM-FM Models, Dual Domain Tracking, Multichannel Tracking, Particle Filter, SIR Filter,

1. INTRODUCTION

High performance target tracking is a critical component of many military weapon delivery systems. For example, precision guided weapons rely on track filters to provide the on-board guidance computer with robust and accurate estimates of the state trajectory of the target in order to successfully maintain the target within the sensor field of view (FOV) and steer the weapon to the target. Here, we focus our attention on the case of a single band imaging infrared (IR) sensor operating in the longwave 8-12 μm (LWIR) band. The sensor delivers a sequence of IR video frames f_k depicting the target against an ambient background scene. Highly maneuverable targets immersed in strong, structured clutter present significant challenges for the tracking system. During a maneuver, the target signature can exhibit *substantial* nonstationary evolution which is difficult or impossible to model accurately by applying spatial transformations to a theoretical appearance model or a small collection of stored library signatures. For any particular target type, even if a *large* collection of views at various poses and magnifications is maintained on-board the weapon system, in real combat one will often encounter observations of the target that fail to match any of the stored signatures well enough to provide reliable target detection and clutter discrimination.

Consider the well-known AMCOM closure sequence [1–4] *rng19_13* depicted in Fig. 2(m)-(p). A closeup view of the lead target from a frame near the one shown in Fig. 2(n) is given in Fig. 1(a). A closeup view of this same target from a frame acquired near the end of the turning maneuver, temporally near the one shown in Fig. 2(o), is given in Fig. 1(b). This is an example of substantial nonstationary target signature evolution that is extremely difficult to model accurately using *a priori* appearance models [4–6] and/or stored signature libraries. Our goal is to develop a new type of track processor that, given an initial target observation (*e.g.*, from a manual designation or a dedicated detection process based on a stored signature library), can maintain a robust track lock without any further need for explicit *a priori* appearance models – even in the presence of severe target signature evolution and severe clutter.

Techniques for tracking based on an initial target observation, or *template* (*e.g.*, matched filter kernel), have been studied extensively. In the presence of highly evolutionary target signatures, the principal difficulty is in identifying when the current template, which is an *empirical* appearance model, has become stale and needs to be refreshed. This is known as the "template update" problem [7–10]. Typical strategies include no update, updating the template every frame, and updating the template every L frames. If the template is updated every frame, we have found that the track filter invariably over adapts, ultimately losing the target and locking onto structured features of the clutter. While the fixed interval update strategy represents an improvement in this regard, it is plagued by the unsolved problem of how to choose the update interval L. In [11, 12], we showed by a proof of concept experiment (using a normalized correlation tracker) that it is possible to detect when the template has become stale by observing the target simultaneously in *two* domains: *viz.,* the pixel domain and the modulation domain.

In the presence of highly evolutionary target signatures, we found that a pixel domain correlation tracker fails frequently, and also that a modulation domain correlation tracker fails frequently. However, our work in [11, 12] suggests *strongly* that the failure modes in the two domains are almost never the same. Thus, even when both domains fail simultaneously, it is extremely rare that the two simultaneous failures are similar. Thus, one can reliably detect a stale template condition by checking for divergence of the pixel domain track centroid and the modulation domain track centroid. In [12], we showed that, with an idealized detection process and a pair of normalized correlation trackers operating in the pixel and modulation domains, 95% of the stale template conditions that arise in the AMCOM IR closure sequences are detectable by thresholding the divergence of the pixel domain and modulation domain track centroids.

To make this notion practical, in this paper we will introduce a new multichannel dual domain SIR track filter capable of successfully maintaining track lock and obtaining a high probability of kill against virtually all of the targets in the LWIR AMCOM closure sequences with absolutely no *a priori* knowledge of any kind after the initial target designation.
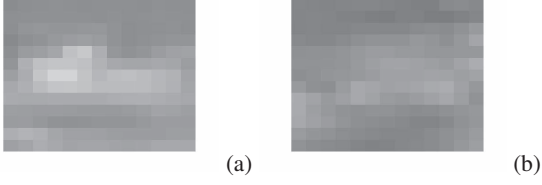
(a)   (b)

Fig. 1. Two radically differing views of the lead target in AMCOM closure sequence *rng19_13*. (a) Appearance before the turning maneuver. (b) Appearance just after the turning maneuver.

2. MULTICHANNEL DUAL DOMAIN TRACK FILTER

We formulate the dual domain track filter in a unified Bayesian framework where three SIR particle filters run in tandem to estimate the target state, including both magnification and position. The main SIR filter models the target jointly in both the pixel and modulation domains and produces the position estimates that are output by the overall tracking system. The remaining two SIR filters model the target in the pixel domain and modulation domain individually; divergence of the tracked position in these two filters indicates a stale template condition. All three SIR filters are initialized with an *a priori* target designation in the first frame of the sequence. This designation, which could be obtained from a dedicated, independent detection process or from a manual designation with a human in the loop, includes the target centroid, height, and width, all specified in units of pixels. The templates in all three SIR filters are refreshed from the last reliably tracked frame when the distance between the expected modulation domain and pixel domain centroids exceeds an empirical threshold. For all of the LWIR AMCOM sequences, we set the threshold to three pixels.

The pixel domain SIR filter inputs raw LWIR frames directly from the sensor, while the remaining SIR filters incorporate modulation domain measurements that must be computed from the pixel domain frames as described in [13]. Each frame f_k is considered to be the real part of a complex-valued image t_k that is modeled as a multi-component AM-FM function according to

$$
\begin{aligned}
t_k(x_1, x_2) &= f_k(x_1, x_2) + j\mathcal{H}[f_k(x_1, x_2)] \\
&= \sum_{m=1}^{M} a_m(x_1, x_2) \exp[j\varphi_m(x_1, x_2)], \quad (1)
\end{aligned}
$$

where $\mathcal{H}[\cdot]$ is the partial Hilbert transform [13, 14]. We characterize the frame f_k in the modulation domain by computing estimates of the M AM functions a_m and the M FM functions $\nabla\varphi_m$ in (1). First, a multichannel Gabor filterbank with M channels having impulse responses g_m is applied to t_k to obtain response images $y_m = t_k * g_m \approx a_m e^{j\varphi_m} * g_m$, $1 \le m \le M$, that are each dominated on a pointwise basis by a single image component in (1). The M FM functions in (1) may then be estimated via [13]

$$
\nabla\varphi_m(x_1, x_2) \approx \text{Re}\left[\frac{\nabla y_m(x_1, x_2)}{j y_m(x_1, x_2)}\right]. \quad (2)
$$

Once the frequency estimates have been obtained using (2), the AM functions in (1) are estimated according to $a_m(x_1, x_2) \approx |y_m(x_1, x_2)/G_m[\nabla\varphi_m(x_1, x_2)]|$, where G_m is the frequency response of the m'th filterbank channel [13]. For the LWIR AMCOM sequences, we apply the same Gabor filterbank that was used in [12], which has $M = 18$ channels.

Although our results in [14] strongly suggest that the full modulation domain characterization is useful for enhancing target-background class separability in IR imagery, for the AMCOM data we find in practice that the AM functions are considerably more powerful that the FM functions in this regard. Thus, throughout this paper we use the estimated FM functions (2) in computing the AM estimates a_m, but only the AM functions are used as modulation domain input to the SIR particle filters. For each frame f_k, the input to the pixel domain SIR filter is the frame f_k itself, the input to the modulation domain filter is a stack of the 18 AM images a_m computed from f_k, and both the frame and the AM images are input to the joint domain SIR filter.

We model the target appearance and kinematics using a six component state vector $\mathbf{x}_k$ defined by

$$
\mathbf{x}_k = [x_{1,k}\ \dot{x}_{1,k}\ \delta_{1,k}\ x_{2,k}\ \dot{x}_{2,k}\ \delta_{2,k}]^T = [\mathbf{x}_{1,k}^T\ \mathbf{x}_{2,k}^T], \quad (3)
$$

where $\mathbf{x}_{1,k} = [x_{1,k}\ \dot{x}_{1,k}\ \delta_{1,k}]^T$ and $\mathbf{x}_{2,k} = [x_{2,k}\ \dot{x}_{2,k}\ \delta_{2,k}]^T$. In (3), $x_{1,k}$ and $x_{2,k}$ are the horizontal and vertical image coordinates of the template spatial centroid, $\dot{x}_{1,k}$ and $\dot{x}_{2,k}$ are the corresponding velocities in pix/frame, and $\delta_{1,k}$ and $\delta_{2,k}$ are the horizontal and vertical extent of the template in pixels. The template has 19 planes of size $\delta_{1,k} \times \delta_{2,k}$ corresponding to the LWIR frame f_k and the 18 AM images a_m. The state vector $\mathbf{x}_k$ indicates that the current target template appears in the image stack $f_k|a_1|\ldots|a_{18}$ at pixel $x_{1,k}, x_{2,k}$ with size $\delta_{1,k} \times \delta_{2,k} \times 19$. The state transition equation is a white noise acceleration model given by

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

where

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

$\mathbf{v}_k = [v_{1,k}\ 0\ 0\ v_{2,k}\ 0\ 0]^T$, $v_{1,k}$ and $v_{2,k}$ are uncorrelated zero-mean Gaussian noises, Δ is the interframe time, and γ is a uniform ternary process that models the frame-to-frame magnification change by

$$
\gamma = \begin{cases} -\alpha, & p = 1/3, \\ 0, & p = 1/3, \\ \beta, & p = 1/3. \end{cases} \quad (6)
$$

In (6), p indicates probability and α and β are magnification gain parameters related to the rate at which the weapon closes on the target. For the AMCOM sequences, we generally set $\alpha = \beta = 0.1$. The observation equation is given by $\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k$, where $\mathbf{n}_k = [n_{1,k}\ n_{2,k}]$, $n_{1,k}$ and $n_{2,k}$ are uncorrelated zero-mean Gaussian noise processes, and

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

Let I_k^i be the target template hypothesized in frame k by the particle $(\mathbf{x}_k^i, w_k^i)$ [15]. We generate I_k^i from the original (or most recently refreshed) template by applying bicubic interpolation to account for the change in size. Let I_k denote the subset of the image stack $f_k|a_1|\ldots|a_{18}$ that is covered by the template I_k^i. Then we define the likelihood function to be $p(\mathbf{z}_k|\mathbf{x}_k^i) \propto \exp[-\text{MSE}(I_k^i, I_k)]$, where $\text{MSE}(I_k^i, I_k)$ is the mean squared error between I_k^i and I_k. The MSE is calculated only over the appropriate planes of the template; *e.g.*, in the pixel domain SIR filter it is calculated only over the plane corresponding to the frame f_k, for the modulation domain filter it is calculated only over the planes corresponding to the AM images a_m, and for the joint domain filter it it calculated over all 19 planes. Details of the SIR filters are as described in [15].

3. RESULTS AND DISCUSSION

We evaluated the performance of the multichannel dual-domain particle filter (DDPF) introduced in Section 2 against the eight LWIR AMCOM closure sequences listed in the first column of Table 1. The experiment was as described in the first paragraph of Section 2: the track filter was initialized with an initial designation including the target centroid and spatial extent. After the initial frame, no further *a priori* information of any kind was provided to the track filter. The average absolute error in the DDPF track centroid relative to ground truth is given in the third column of Table 1 in pixel units. Illustrative examples depicting key frames from several of the sequences are shown in Fig. 2, where the spatial extent of the template (*e.g.*, the *track gate*) is superimposed on the raw pixel domain imagery. The first row of Fig. 2 shows key frames from the visible wavelength test sequence *hand-ball*, which we used to develop and tune the ternary magnification change model (4)-(6). The remaining rows of Fig. 2 illustrate the performance of the DDPF against LWIR AMCOM closure sequences. In every case, the DDPF tracked the targets with sufficient accuracy to provide a high probability of kill at the end of the sequence. These sequences are notoriously difficult because the targets generally exhibit substantial magnification and pose changes as well as severe nonstationary signature evolution. In our opinion, the lead vehicle in the sequence *rng19_13* (fourth row of Fig. 2; also see Fig. 1) represents one of the most difficult challenges in the entire AMCOM data set.

For comparison, we also ran a pure modulation domain SIR particle filter (MDPF) and a pure pixel domain SIR particle filter (PDPF) against the sequences listed in Table 1. In every case, both the MDPF and the PDPF failed to maintain track lock and lost the target. This was true regardless of the template update strategy (*i.e.*, no update, update every frame, fixed interval update).

To gain additional insight into the performance gains that are possible with multichannel dual domain measurements, we supplied both the MDPF and the PDPF with the dual domain track consistency checks from the DDPF. Thus, in this modified experiment, the templates for all three filters were updated at the same times using dual domain information. The resulting centroid errors for the MDPF and PDPF are given in the last two columns of Table 1. The bottom row of the table gives the percent improvement in the absolute centroid error relative to the PDPF with dual domain updates. These data demonstrate that, in addition to making it possible to detect stale templates and maintain track lock, the multichannel dual domain measurements also provide an inherent accuracy advantage compared to traditional pixel domain techniques.

4. REFERENCES

[1] J.F. Khan and M.S. Alam, "Efficient target detection in cluttered FLIR imagery," in *Optical Pattern Recognition XVI*, D.P. Casasent and T.-H. Chao, Eds., 2005, vol. 5816 of *Proc. SPIE*, pp. 39–53.

[2] S. Yi and L. Zhang, "A novel multiple tracking system for UAV platforms," in *ISR Systems and Applications III*, D.J. Henry, Ed., 2006, vol. 6209 of *Proc. SPIE*.

[3] A. Dawoud, M.S. Alam, A. Bai, and C. Loo, "Decision fusion algorithm for target tracking in infrared imagery," *Opt. Eng.*, vol. 44, no. 2, 2005.

[4] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, pp. 1–45, Dec. 2006.

Table 1. Average absolute error in tracked position (pix). All three trackers used dual domain consistency checks for template update.

Sequence	Frames	DDPF	MDPF	PDPF
rng14_15	135	**2.858**	3.512	3.645
rng15_20	124	**1.815**	2.092	2.030
rng16_18	135	**2.012**	2.153	2.142
rng18_16	179	**1.234**	1.583	1.818
rng19_06	165	**2.645**	2.760	2.888
rng19_07	63	**2.548**	2.833	3.039
rng19_13	232	**4.387**	4.704	4.724
rng19_NS	167	**2.283**	2.850	2.926
Improvement		**14.77%**	3.12%	0%

[5] M.G.S. Bruno, "Bayesian methods for multispectral target tracking in image sequences," *IEEE Trans. Image Proc.*, vol. 52, no. 7, pp. 1848–1861, Jul 2004.

[6] M.G.S. Bruno, "Sequential importance sampling filtering for target tracking in image sequences," *IEEE Signal Process. Let.*, vol. 10, no. 8, pp. 246–249, Aug 2003.

[7] I. Matthews, T. Ishikawa, and S. Baker, "The template update problem," *IEEE Trans. Pattern Anal., Machine Intel.*, vol. 26, no. 6, pp. 810–815, Jun. 2004.

[8] L.J. Latecki and R. Miezianko, "Object tracking with dynamic template update and occlusion detection," in *Proc. IEEE Int'l. Conf. Pattern Recog.*, Hong Kong, China, Aug. 20-24, 2006, vol. 1, pp. 556–560.

[9] G.D. Harger and P.N. Belhumeur, "Real-time tracking of image regions with changes in geometry and illumination," in *Proc. IEEE Int'l. Conf. Computer Vision, Pattern Recog.*, San Francisco, CA, Jun. 18-20 1996, pp. 403–410.

[10] Z. Peng, Q. Zhang, and A. Guan, "Extended target tracking using projection curves and matching pel count," *Optical Eng.*, vol. 46, no. 6, pp. 066401–1 – 066401–6, Jun. 2007.

[11] N.A. Mould, C.T. Nguyen, and J.P. Havlicek, "Infrared target tracking with AM-FM consistency checks," in *Proc. IEEE Southwest Symp. Image Anal., Interp.*, Santa Fe, NM, Mar. 24-26, 2008, pp. 5–8.

[12] N.A. Mould, C.T. Nguyen, C.M. Johnston, and J.P. Havlicek, "Online consistency checking for AM-FM target tracks," in *Proc. SPIE/IS&T Conf. Computational Imaging VI*, C.A. Bouman, E.L. Miller, and I. Pollak, Eds., 2008, vol. 6814 of *Proc. SPIE*, pp. 681413–1 – 681413–12.

[13] J.P. Havlicek, D.S. Harding, and A.C. Bovik, "Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models," *IEEE Trans. Image Proc.*, vol. 9, no. 2, pp. 227–242, Feb. 2000.

[14] J.P. Havlicek, C.T. Nguyen, and M. Yeary, "Modulation domain infrared target models," in *Targets and Backgrounds XII: Characterization and Representation*, W. R. Watkins and D. Clement, Eds., 2006, vol. 6239 of *Proc. SPIE*, pp. 62390D–1 – 62390D–11.

[15] 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.*, vol. 50, no. 2, pp. 174–188, 2002.
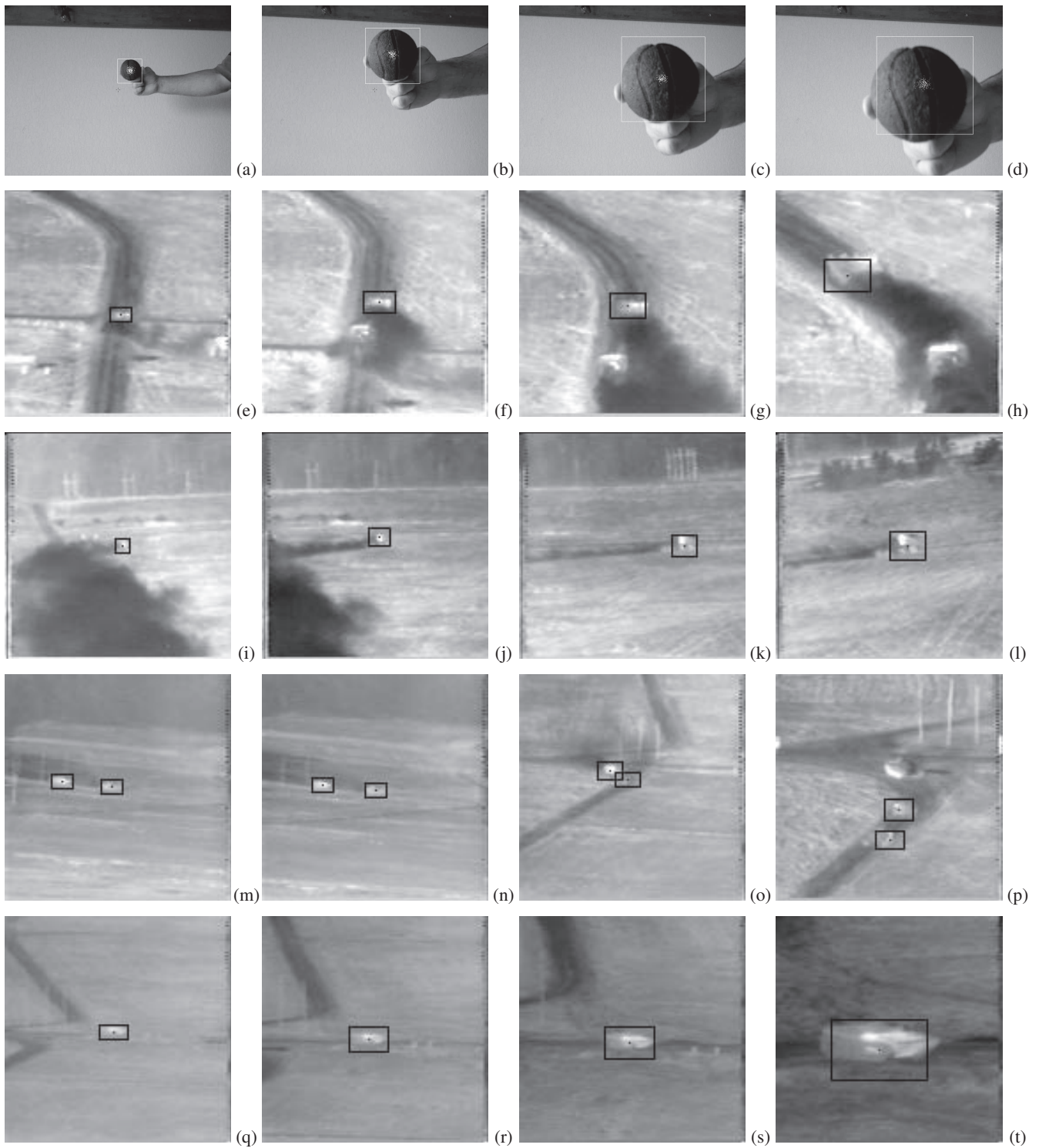
Fig. 2. Multichannel dual domain tracking examples. (a)-(d): *hand-ball* sequence (with track gate superimposed): (a) Frame 1. (b) Frame 117. (c) Frame 149. (d) Frame 160. (e)-(t): LWIR AMCOM closure sequence examples (with track gate superimposed). (e)-(h): sequence *rng19_NS*, frames (e) 208, (f) 228, (g) 244, and (h) 270. (i)-(l): sequence *rng19_07*, frames (i) 132, (j) 147, (k) 169, and (l) 194. (m)-(p): sequence *rng19_13*, frames (m) 49, (n) 61, (o) 168, and (p) 230. (q)-(t): sequence *rng16_18*, frames (q) 140, (r) 184, (s) 222, and (t) 273.