Modulation Domain Template Tracking

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

{chuong,joebob,yeary}@ou.edu

Abstract

For the first time, we perform normalized correlation template tracking in the modulation domain. For each frame of the video sequence, we compute a multi-component AM-FM image model that characterizes the local texture structure of objects and backgrounds. Tracking is carried out by formulating a modulation domain correlation function in the derived feature space. Using visible and long-wave infrared sequences as illustrative examples, we study the performance of this new approach relative to two basic pixel domain correlation template trackers. We also present preliminary results from a new dual domain tracker that operates simultaneously in both the pixel and modulation domains.

1. Introduction

Although computationally expensive, template tracking is of interest in applications where well known techniques such as the Kalman filter, extended Kalman filter (EKF), and particle filter cannot be conveniently formulated due to insufficient *a priori* knowledge of the object kinematics that are expected to be observed in the image plane [1, 16, 18]. Situations of this type arise, for example, in military applications where the object signatures and dynamics may be poorly known or even totally unknown prior to the initiation of tracking, where the sensor platform may be subject to strong ego-motion relative to the imaged objects and backgrounds, and where it is infeasible to compensate for the ego-motion in real-time via registration techniques such as those given in [13, 16] due to insufficient computational bandwidth and/or insufficient inertial measurements. Template tracking is commonly applied directly to the video frames acquired from an imaging sensor, where the peak of the normalized correlation between the template and local neighborhoods of the most recently acquired frame is

declared as the track centroid (see, *e.g.*, the recent survey of tracking techniques given in [17]). We will refer to techniques of this type as *pixel domain* template tracking.

In many applications, including notably those where an imaging electro-optical sensor is used, there may be highly structured clutter and backgrounds exhibiting signatures that are as strong as or stronger than the signatures of the objects of interest. Hence, the object-background class separability that can be achieved by processing the raw pixel domain data alone is typically poor and pixel domain template tracking is subject to failure. In such cases, a signal processing advantage can be obtained by performing object detection and tracking in a transformed space or in a joint feature space [4, 12]. Based on preliminary results appearing in [11], it has been suggested recently that object-background class separability in midwave and longwave infrared imagery may be tangibly enhanced by performing signal processing in the modulation domain. However, modulation domain target tracking has not been studied previously. In this paper, we formulate a normalized correlation template tracker in the modulation domain for the first time. We study the performance of this new technique relative to elementary pixel domain template trackers with fixed and updated templates, where visible infrared video sequences are used as illustrative examples.

Because this work is exploratory in nature and because our main goal is to introduce the modulation domain template tracking approach and demonstrate its feasibility, we limit our attention to the case of simple detection processing. In particular, we consider only static templates and temporally updated templates of a fixed size and shape. This restriction is unrealistic for practical deployment systems that must typically deal with strong evolution of the object pose and magnification, as well as partial and full occlusions. However, our results demonstrate that modulation domain template tracking is both feasible and capable of delivering good tracking performance. We also briefly examine a preliminary strategy for performing template tracking jointly in the pixel and modulation domains and obtain a performance that is substantially superior to tracking in ei-

ther domain individually. In Section 2, we briefly review the basic pixel domain normalized correlation template tracker, while the modulation domain normalized correlation function is formulated in Section 3. We compare the performance of the pixel domain, modulation domain, and joint domain trackers in Section 4. For the experiments we assume that a target detection is available in the first frame to initialize the trackers, either from an auxiliary detection process or from a manual designation.

2. Pixel domain template tracking

In this section, we formulate a simple, well-known strategy for performing template tracking in the pixel domain (see, *e.g.*, [15]). We assume that a temporally evolving scene composed of a moving object immersed in a structured background is observed by an imaging sensor. The sensor delivers a sequence of video frames $s(\mathbf{m}, k)$, where $\mathbf{m} \in \mathbb{Z}^2$ is the discrete spatial position vector and $k \in \mathbb{Z}$ is the discrete time index. We assume that the spatial signature of the target is unknown *a priori*, but that a template of size $(2K+1) \times (2L+1)$ pixels that is matched to the object signature in the initial frame $s(\mathbf{m}, k_0)$ is available from an auxiliary detection process or a manual designation. The template is denoted $T(\mathbf{u})$, where $\mathbf{u} \in \mathcal{O} = [-K, K] \times [-L, L]$.

The normalized cross correlation between the template $T(\mathbf{u})$ and a $(2K+1) \times (2L+1)$ neighborhood of the frame $s(\mathbf{m}, k)$ centered about the pixel $\mathbf{m}$ is given by

$$\gamma_{\text{pix}}(\mathbf{m}, k) =$$
$$\sum_{\mathbf{u} \in \mathcal{O}} \left[s(\mathbf{m} + \mathbf{u}, k) - \langle s(\mathbf{m}, k) \rangle \right] \left[T(\mathbf{u}) - \langle T \rangle \right]$$
$$\times \left[\sum_{\mathbf{u} \in \mathcal{O}} \left[s(\mathbf{m} + \mathbf{u}, k) - \langle s(\mathbf{m}, k) \rangle \right]^2 \right]^{-\frac{1}{2}}$$
$$\times \left[\sum_{\mathbf{u} \in \mathcal{O}} \left[T(\mathbf{u}) - \langle T \rangle \right]^2 \right]^{-\frac{1}{2}}, \qquad (1)$$

where $\langle T \rangle = |\mathcal{O}|^{-1} \sum_{\mathbf{u} \in \mathcal{O}} T(\mathbf{u})$ and $\langle s(\mathbf{m}, k) \rangle = |\mathcal{O}|^{-1} \sum_{\mathbf{u} \in \mathcal{O}} s(\mathbf{m} + \mathbf{u}, k)$ are the mean values of the template and of the $(2K+1) \times (2L+1)$ neighborhood of $s(\mathbf{m}, k)$ centered about pixel $\mathbf{m}$.

The pixel domain track centroid for frame k is given by

$$\mathcal{C}_{\text{pix}}(k) = \underset{\mathbf{m}}{\arg\max}\, \gamma_{\text{pix}}(\mathbf{m}, k), \qquad (2)$$

which must in general be evaluated by exhaustive search. If reasonable bounds can be placed on the observed velocity of the object based on *a priori* knowledge of the geometry and of the kinematics of both the object and the sensor platform, then it may be possible to reduce the complexity of (2) by limiting the search to a small neighborhood about $\mathcal{C}_{\text{pix}}(k-1)$ or by applying the computationally efficient gradient descent algorithm given in [9].

3. Modulation domain template tracking

A multi-component AM-FM image model is a dense feature space where the feature vectors consist of spatially local amplitude and frequency modulations that characterize the local texture structure of the image [2, 3, 6]. When signal processing is applied directly to these feature vectors, it is referred to as processing in the *modulation domain* [7]. To perform modulation domain target tracking, we first compute a multi-component AM-FM model for each video frame $s(\mathbf{m}, k)$ and then compute a normalized correlation function between the AM-FM model and a modulation domain object template. Computation of the model and formulation of the correlation function are described below in Sections 3.1 and 3.2.

3.1. AM-FM image model

In this section, we briefly review the technique given in [6] for computing multi-component AM-FM image models. Assume that the digital video frame $s(\mathbf{m}, k)$ contains samples of a spatially continuous frame $s(\mathbf{x}, k)$, where $\mathbf{x} \in \mathbb{R}^2$. In the interest of brevity and clarity, we discuss demodulation with reference to the continuous frame $s(\mathbf{x}, k)$; for a practical implementation the demodulation algorithms stated here should be appropriately discretized as described in [6] or [10]. So that the instantaneous amplitude and frequency of $s(\mathbf{x}, k)$ can be well defined, let $z(\mathbf{x}, k) = s(\mathbf{x}, k) + j\mathcal{H}[s(\mathbf{x}, k)]$ be a complex extension of $s(\mathbf{x}, k)$ computed using the partial 2-D Hilbert transform [3, 5, 6].

The AM-FM image model considers that $z(\mathbf{x}, k)$ is a sum of components of the form

$$z_p(\mathbf{x}, k) = A_p(\mathbf{x}, k) \exp[j\varphi_p(\mathbf{x}, k)], \qquad (3)$$

so that

$$z(\mathbf{x}, k) = \sum_{p=1}^{P} z_p(\mathbf{x}, k) = \sum_{p=1}^{P} A_p(\mathbf{x}, k) \exp[j\varphi_p(\mathbf{x}, k)].$$
$$(4)$$

For each component $z_p(\mathbf{x}, k)$, the AM function $A_p(\mathbf{x}, k)$ characterizes the local texture contrast while the FM function $\nabla\varphi_p(\mathbf{x}, k)$ characterizes the local texture orientation and granularity.

If the component $z_p(\mathbf{x}, k)$ could be observed directly, then the AM and FM functions could be obtained according to [6]

$$\nabla\varphi_p(\mathbf{x}, k) = \text{Re}\left[\frac{\nabla z_p(\mathbf{x}, k)}{j z_p(\mathbf{x}, k)} \right] \qquad (5)$$

and

$$A_p(\mathbf{x}, k) = |z_p(\mathbf{x}, k)|, \qquad (6)$$

which may be verified by direct calculation using (3). In practice, however, only the image $z(\mathbf{x}, k)$ is observable, and

some type of filtering must by applied to isolate the components $z_p(\mathbf{x}, k)$. Following [2, 3, 10], we process the frame $z(\mathbf{x}, k)$ with a bank of P circularly symmetric Gabor filters arranged to tile the right frequency half-plane, where each filter has a half-peak radial bandwidth of one octave. Let $g_p(\mathbf{x})$ and $G_p(\boldsymbol{\omega})$ be the impulse response and frequency response of the p'th filter. Under the assumption that the response is given by

$$y_p(\mathbf{x}, k) = z(\mathbf{x}, k) * g_p(\mathbf{x}) \approx z_p(\mathbf{x}, k) * g_p(\mathbf{x}), \quad (7)$$

it may be shown that

$$\nabla \varphi_p(\mathbf{x}, k) \approx \mathrm{Re} \left[\frac{\nabla y_p(\mathbf{x}, k)}{j y_p(\mathbf{x}, k)} \right] \quad (8)$$

and

$$A_p(\mathbf{x}, k) \approx \left| \frac{y_p(\mathbf{x}, k)}{G_p \left[\nabla \varphi_p(\mathbf{x}, k) \right]} \right| \quad (9)$$

with an error that was bounded in [6]. The modulation domain feature vectors are given by $[A_p(\mathbf{x}, k)\, R_p(\mathbf{x}, k)\, \theta_p(\mathbf{x}, k)]^T$, where

$$\theta_p(\mathbf{x}, k) = \arg \nabla \varphi_p(\mathbf{x}, k) \quad (10)$$

is the local texture orientation and $R_p(\mathbf{x}, k) = |\nabla \varphi_p(\mathbf{x}, k)|$ characterizes the local texture granularity. In a discrete implementation, one obtains P feature vectors $[A_p(\mathbf{m}, k)\, R_p(\mathbf{m}, k)\, \theta_p(\mathbf{m}, k)]^T$ at each pixel $\mathbf{m}$, where P is the number of components in the model (4) and is by defined by number of filterbank channels.

3.2. Modulation domain correlation function

In Section 2, we assumed that the object signature (template) $T(\mathbf{u})$ in the initial frame $s(\mathbf{m}, k_0)$ was known from an auxiliary detection process or a manual designation, *i.e.*, that $T(\mathbf{u})$ was given by a particular $(2K + 1) \times (2L + 1)$ neighborhood in the frame $s(\mathbf{m}, k_0)$ centered about some pixel $\mathbf{m}_0$. By extracting corresponding neighborhoods of size $(2K + 1) \times (2L + 1)$ centered about the pixel $\mathbf{m}_0$ from the modulation domain images $A_p(\mathbf{m}, k_0)$, $R_p(\mathbf{m}, k_0)$, and $\theta_p(\mathbf{m}, k_0)$, we obtain a set of $3P$ templates $T_p^A(\mathbf{m})$, $T_p^R(\mathbf{m})$, and $T_p^\theta(\mathbf{m})$ that are matched to the modulation domain signature of the object in frame $s(\mathbf{m}, k_0)$.

Intuitively, we define the modulation domain normalized correlation function $\gamma_{\mathrm{mod}}(\mathbf{m}, k)$ as an equally weighted average of the normalized correlation function between each of these templates and the corresponding components of the feature vectors $[A_p(\mathbf{m}, k)\, R_p(\mathbf{m}, k)\, \theta_p(\mathbf{m}, k)]^T$ computed from the frame $s(\mathbf{m}, k)$. Let $\mathcal{O}$ be defined as in Section 2. Formally, we take

$$\gamma_{\mathrm{mod}}(\mathbf{m}, k) = \frac{\sum_{p=1}^{P} \gamma_p^A(\mathbf{m}, k) + \gamma_p^R(\mathbf{m}, k) + \gamma_p^\theta(\mathbf{m}, k)}{3P}, \quad (11)$$

where

$$\gamma_p^A(\mathbf{m}, k) =$$
$$\sum_{\mathbf{u} \in \mathcal{O}} \left[A_p(\mathbf{m} + \mathbf{u}, k) - \langle A_p(\mathbf{m}, k) \rangle \right] \left[T_p^A(\mathbf{u}) - \langle T_p^A \rangle \right]$$
$$\times \left[\sum_{\mathbf{u} \in \mathcal{O}} \left[A_p(\mathbf{m} + \mathbf{u}, k) - \langle A_p(\mathbf{m}, k) \rangle \right]^2 \right]^{-\frac{1}{2}}$$
$$\times \left[\sum_{\mathbf{u} \in \mathcal{O}} \left[T_p^A(\mathbf{u}) - \langle T_p^A \rangle \right]^2 \right]^{-\frac{1}{2}}, \quad (12)$$

and where $\langle A_p(\mathbf{m}, k) \rangle = |\mathcal{O}|^{-1} \sum_{\mathbf{u} \in \mathcal{O}} A_p(\mathbf{m}, k)$ and $\langle T_p^A(\mathbf{u}) \rangle = |\mathcal{O}|^{-1} \sum_{\mathbf{u} \in \mathcal{O}} T_p^A(\mathbf{u})$. The terms $\gamma_p^R(\mathbf{m}, k)$ and $\gamma_p^\theta(\mathbf{m}, k)$ are defined analogously to (12).

The modulation domain track centroid for frame k is then given by

$$\mathcal{C}_{\mathrm{mod}}(k) = \underset{\mathbf{m}}{\mathrm{argmax}}\, \gamma_{\mathrm{mod}}(\mathbf{m}, k). \quad (13)$$

As with (1), the naïve, brute force approach is to evaluate (13) by exhaustive search. If *a priori* information about the geometry and the kinematics of the sensor platform and the objects of interest are available, then they can be used to limit the search space to a neighborhood about $\mathcal{C}_{\mathrm{mod}}(k-1)$. Alternatively, if the sensor ego-motion is not too severe, then it may be possible to use a Kalman filter or EKF to predict the track centroid based on recent observations and limit the search to a neighborhood (or track gate) about the predicted location.

4. Examples

In this section, we illustrate the modulation domain normalized correlation template tracker developed in Section 3 and characterize its performance relative to comparable pixel domain techniques. In all cases, the object to be tracked is designated in the initial frame and the designated signature is used as the initial template. The shape and spatial extent of the template are not adapted. In both the pixel domain and modulation domain, we consider two elementary cases. In the first case, denoted *static*, the initial template is used throughout the entire sequence. In the second case, denoted *updated*, the template is updated every frame using the observed object signature from the previous frame. The frame size in all of the examples is 128×128 pixels, and the Gabor filterbank comprises 18 channels ($P = 18$).

In Fig. 1, the object of interest is the face of a human who is walking through the field of view. The sensor is a visible wavelength digital video camera and the length of the sequence is 40 frames. Each row of the figure corresponds to a video frame, where the frame number is denoted at the far left. The raw video frames are shown in

Figure 1. Modulation domain and pixel domain template tracking with static and updated templates for the task of tracking a human face in a visible wavelength video sequence. Results for the four trackers are shown in columns 1-4, while the modulation domain correlation function is depicted in the rightmost column.
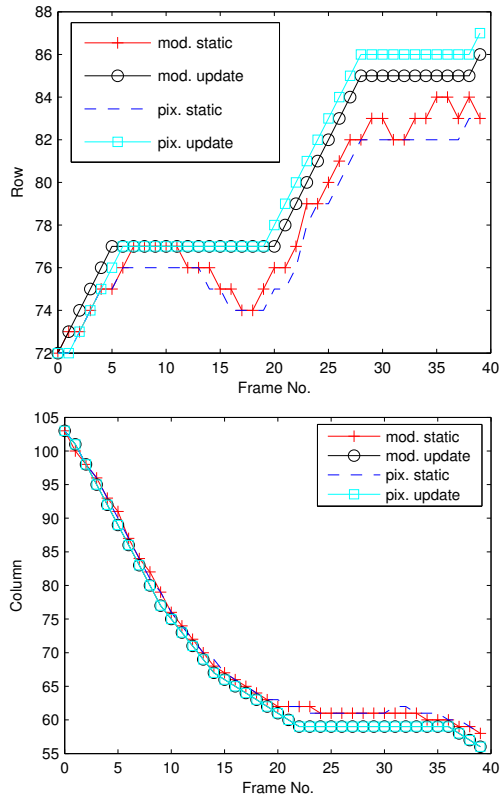
Figure 2. Row and column of the track centroid for all four methods as function of the frame number for the example in Fig. 1.

Frames Dropped (out of 150)				
pixel domain		modulation domain		dual
static	updated	static	updated	domain
32	11	72	0	0

Table 1. Tracking performance on the example of Fig. 3 measured in terms of the number of dropped frames where track lock could not be maintained.

the first column. The track centroids for the pixel domain tracker with static and updated templates and the modulation domain tracker with static template are overlaid on the frames in columns two through four, where the tracking method is denoted along the top of the figure. Although the centroids are not shown for the modulation domain tracker with updated templates, they are not significantly different visually from the result for the static modulation domain template. The rightmost column depicts the modulation domain normalized correlation function. Despite the presence of some structured background, all four techniques perform well. The upper and lower graphs in Fig. 2 show, respectively, the row and column of the track centroids in units of pixels for all four trackers as a function of frame number. Despite the simplicity of the normalized correlation tracker,

it is a powerful albeit computationally expensive approach. For example, the centroids shown in Fig. 1 exhibit both improved accuracy and reduced jitter compared to a condensation filter and a particle filter where state space partitioning and a bank of EKF's were used to improve the proposal distribution.

Fig. 3 shows tracking results for one of the well known AMCOM infrared missile closure sequences (see, *e.g.*, [1, 8, 14, 18]). With reference to the AMCOM data set, this is longwave sequence `rng18_03`. Although the difficulty of this sequence is typical for the AMCOM data set, it is significantly more challenging than the example in Fig. 1, mainly due to a dim target, ambient noise, and a poor target-to-clutter ratio. The layout of Fig. 3 is identical to that of Fig. 1, where each row corresponds to a frame and each column corresponds to a tracking method. The length of the original sequence was 450 frames. Tracking was initiated in frame 290 and carried out for 150 frames (over this range of frames, the magnification changes in the target are relatively benign compared to the sequence as a whole).

In this case, the static template trackers in both domains fail to track the target, while the updated templates perform relatively much better. The modulation domain template tracker with static template achieves the best performance out of the four basic methods for this sequence. The number of frames dropped by each tracker (out of 150) is shown in Table 1, where a *dropped frame* is one for which the error in the tracked centroid relative to ground truth exceeds the spatial support of the template. For the basic techniques considered in this paper, where pose and scale are not adaptively estimated, the performance seen in this example is fairly typical for the AMCOM data set. It is is not uncommon for either the static or updated approaches to fail. Moreover, both approaches fail on some of the more difficult sequences such as `rng19_13`.

It is interesting to note in columns two and four of Fig. 3 that the pixel domain tracker with static template and the modulation domain tracker with static template seem to exhibit failure modes that are quite different from one another. In a preliminary analysis of 14 of the AMCOM sequences, this seems to occur consistently: the pixel domain and modulation domain trackers (with both static and updated templates) all fail frequently; however, it is rare for failure to occur in both domains on a single frame. This suggests that a significant performance gain might be obtained by tracking jointly in the two domains.

As a preliminary investigation of this idea, we constructed a dual domain track filter where the pixel domain and modulation domain template trackers operate independently with static templates. However, when the distance between the two centroids exceeds an empirical threshold (four pixels in this case), both trackers refresh their template using the target signature observed in the most recently

tracked frame. The results of this approach for `rng18_03` are shown in Fig. 4, along with the modulation domain correlation function. For reference, the track centroids obtained from the pixel domain tracker with updated template as shown in the third column of Fig. 3 are repeated in the first column of Fig. 4. As can be seen from the figure, performance of the dual domain tracker is superior to any of the results given in Fig. 3. This outcome has occurred consistently on the 14 AMCOM sequences on which we have run the dual domain tracker to date.

5. Conclusion

For the first time, we formulated a normalized correlation template tracker in the modulation domain, where object detection and tracking are based on an explicit, spatially and spectrally local texture model. We demonstrated that modulation domain object tracking is feasible and capable of delivering performance comparable to that of traditional pixel domain template trackers. The multiband modulation domain template and correlation function given in Section 3 are distinct from any prior AM-FM or modulation domain technique of which we are aware. We focused on the fundamental ideas behind this new approach and provided a baseline performance comparison with comparable elementary pixel domain techniques. However, we did not consider practical issues including template resizing, pose estimation, scale estimation, and occlusion; these are important topics for future research. Both the modulation domain and pixel domain template trackers performed well against the task of tracking the face of a walking human in a visible wavelength video sequence. Tracking errors in the much more difficult task of tracking a ground vehicle in a longwave infrared AMCOM missile closure sequence were larger for both techniques, but still comparable. For this task, it is very likely that performance in both domains can be dramatically improved by incorporating standard techniques for adaptively resizing the template and estimating pose and scale. Our most significant finding was that, even with the elementary template updating considered here, performance was significantly improved by tracking jointly in both domains. This result has been demonstrated consistently on 14 of the AMCOM closure sequences to date.

References

[1] A. Bal and M. S. Alam. Automatic target tracking in FLIR image sequences using intensity variation function and template modeling. *IEEE Trans. Instrum., Meas.*, 54(5):1846–1852, Oct. 2005.

[2] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE. Trans. Pattern Anal. Machine Intell.*, 12(1):55–73, Jan. 1990.

[3] A. C. Bovik, N. Gopal, T. Emmoth, and A. Restrepo. Localized measurement of emergent image frequencies by Gabor wavelets. *IEEE Trans. Info. Theory*, 38(2):691–712, Mar. 1992.

[4] M. Diani, A. Baldacci, and G. Corsini. Novel background removal algorithm for navy infrared search and track systems. *Optical Eng.*, 40(8):1729–1734, Aug. 2001.

[5] S. L. Hahn. *Hilbert Transforms in Signal Processing*. Artech House, Norwood, MA, 1996.

[6] 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):227–241, Feb. 2000.

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

[8] J. F. Khan, M. S. Alam, R. R. Adhami, and S. M. A. Bhuiyan. Minimax distance transform correlation filter based target detection in FLIR imagery. In A. G. Tescher, editor, *Applications of Digital Image Processing XXVIII*, volume 5909 of *Proc. SPIE*, pages 590907–1 – 590907–12, 2005.

[9] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. IEEE Int'l. Joint Conf. Artificial Intell.*, pages 674–679, Aug. 24-28 1981.

[10] P. Maragos and A. C. Bovik. Image demodulation using multidimensional energy separation. *J. Opt. Soc. Amer. A*, 12(9):1867–1876, Sep. 1995.

[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] B. Porat and B. Friedlander. A frequency domain algorithm for multiframe detection and estimation of dim targets. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(4):398–401, Apr. 1990.

[13] H. Shekarforoush and R. Chellappa. A multi-fractal formalism for stabilization, object detection and tracking in flir sequences. In *Proc. IEEE Int'l. Conf. Image Proc.*, pages 78–81, Vancouver, Canada, Sep. 10-13, 2000, vol. 3.

[14] A. Srivastava, S. H. Joshi, W. Mio, and X. Liu. Statistical shape analysis: clustering, learning, and testing. *IEEE Trans. Pattern Anal. Machine Intell.*, 27(4):590–602, Apr. 2005.

[15] 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.

[16] A. Strehl and J. K. Aggarwal. Detecting moving objects in airborne forward looking infra-red sequences. In *Proc. IEEE Workshop Comp. Vision Beyon. Visi. Spectr.*, pages 3–12, Fort Collins, CO, Jun. 21-22 1999.

[17] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Comput. Surv.*, 38(4):1–45, Dec. 2006. Article 13.

[18] A. Yilmaz, K. Shafique, and M. Shah. Target tracking in airborne forward looking infrared imagery. *Image and Vision Computing*, 21(7):623–635, 2003.
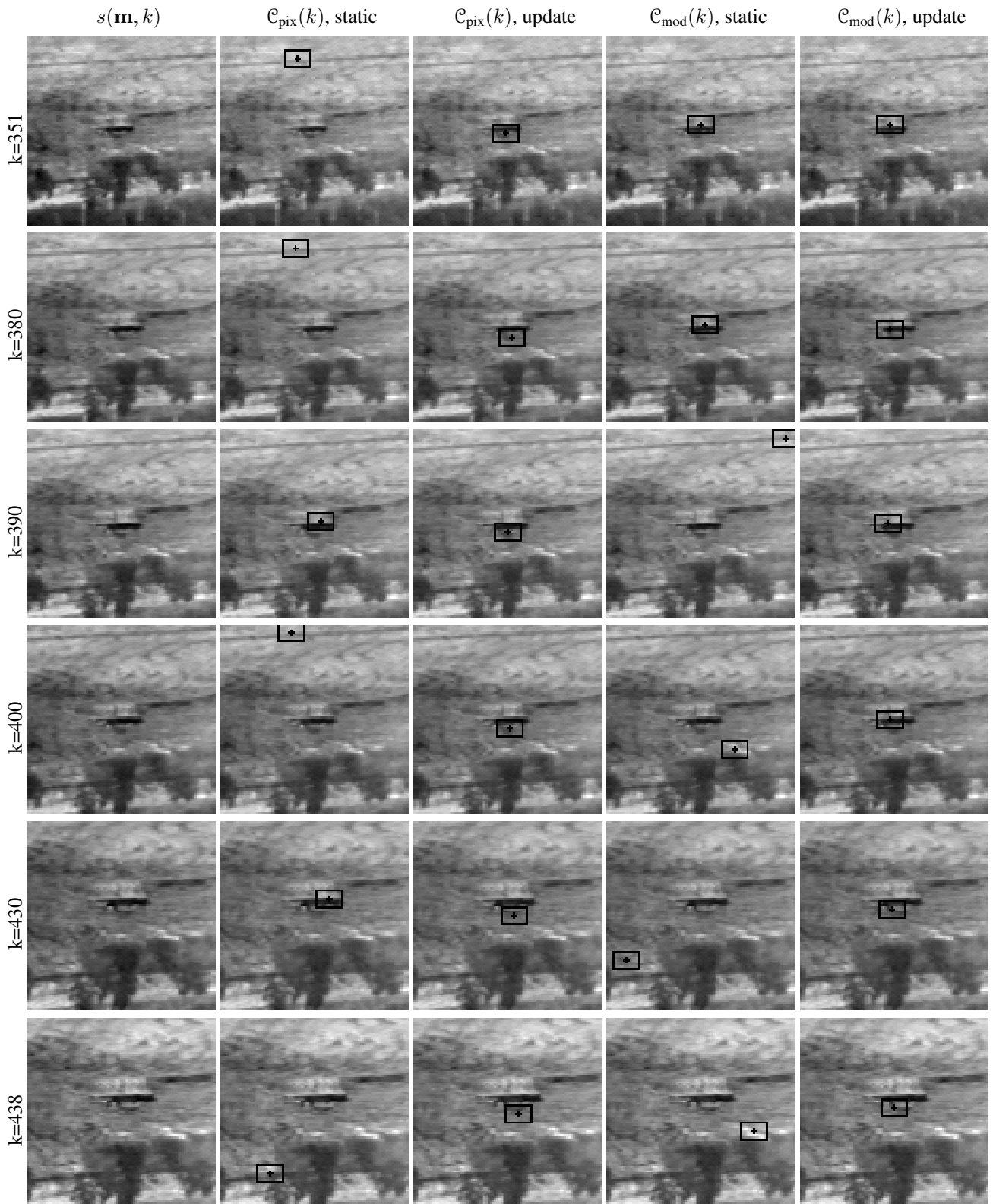
Figure 3. Modulation domain and pixel domain template tracking with static and updated templates for the task of tracking a ground vehicle in a longwave infrared AMCOM missile closure sequence (rng18_03). The basic layout of the figure is the same as in Fig. 1.
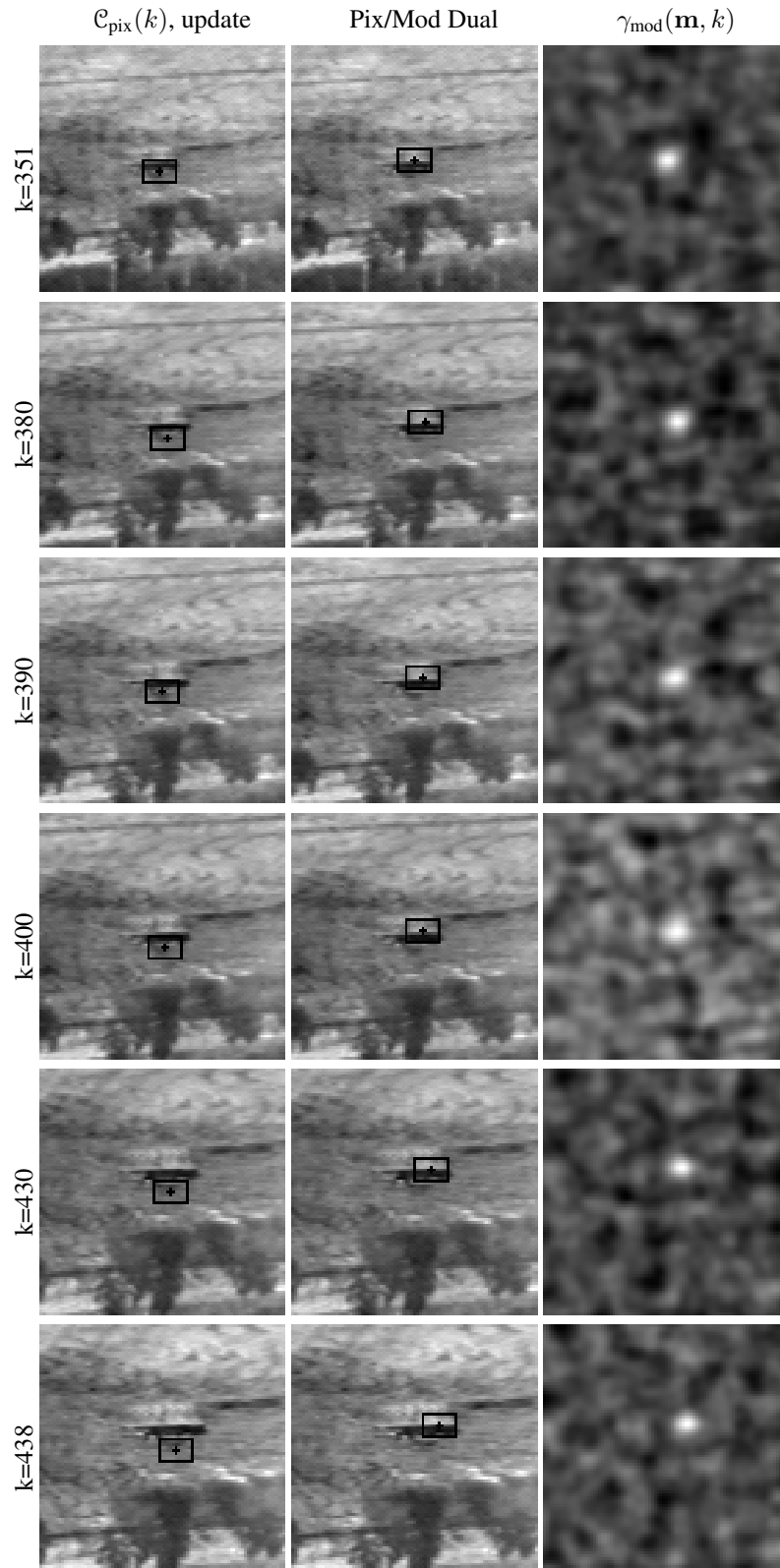
Figure 4. Comparison of dual domain tracking and pixel domain tracking with updated template on AMCOM sequence `rng18_03`. This is the same sequence that was depicted in Fig. 3. The modulation domain correlation function is also shown.