

Online Consistency Checking for AM-FM Target Tracks

Nick A. Mould, Chuong T. Nguyen, Colin M. Johnston, and Joseph P. Havlicek

University of Oklahoma, Norman, OK, USA

ABSTRACT

We compute AM-FM models for infrared video frames depicting military targets immersed in structured clutter backgrounds. We show that independent correlation based detection processes can be implemented in the pixel and modulation domains and used to construct useful online track consistency checks that indicate when the detection process has been degraded due to nonstationary evolution of the target signature. Throughout the paper, we use the well-known AMCOM closure sequences as exemplars.

Keywords: AMCOM, AM-FM, Modulation Domain, Gabor Filter, Target Tracking

1. INTRODUCTION

In this paper, we compute AM-FM models for infrared video frames depicting military targets immersed in structured clutter backgrounds. Throughout the paper, we use the well-known AMCOM closure sequences as exemplars. We show that independent correlation based detection processes can be implemented in the pixel and modulation domains and used to construct useful online track consistency checks that indicate when the detection process has been degraded due to nonstationary evolution of the target signature. The 2-D Fourier transform $S(\boldsymbol{\Omega}) = \int_{\mathbb{R}^2} s(\mathbf{x})e^{-j\boldsymbol{\Omega}^T \mathbf{x}} d\mathbf{x}$ and inverse $s(\mathbf{x}) = (2\pi)^{-2} \int_{\mathbb{R}^2} S(\boldsymbol{\Omega})e^{j\boldsymbol{\Omega}^T \mathbf{x}} d\boldsymbol{\Omega}$ provide a means for representing an image as a sum of sinusoidal components that each has constant amplitude, linear phase, and constant frequency throughout the entire image domain. With the Fourier representation, nonstationary image structure is created by constructive and destructive interference between stationary Fourier components. By contrast, AM-FM image models facilitate the representation of an image in terms of oscillating functions admitting smooth but nonstationary amplitude and frequency modulations.

2. MODULATION DOMAIN SIGNAL MODEL

While the theory of AM-FM signal modeling^{1,2} extends naturally to multiple dimensions, we restrict our attention here to the case of 2-D images. Let $\mathbf{x} = [x_1 \ x_2]^T \in \mathbb{R}^2$ and let $s(\mathbf{x}): \mathbb{R}^2 \rightarrow \mathbb{R}$ be a real-valued optical image (with continuous domain). For the moment, assume that $s(\mathbf{x})$ is locally smooth in the sense that it can be well approximated by a pure sinusoid over any sufficiently small neighborhood. Such images are described as *locally coherent*.^{2,3} We write

$$s(\mathbf{x}) = a(\mathbf{x}) \cos[\varphi(\mathbf{x})], \quad (1)$$

where $a(\mathbf{x}): \mathbb{R}^2 \rightarrow \mathbb{R}^+$ is the amplitude modulation (AM) function of $s(\mathbf{x})$ and $\nabla\varphi(\mathbf{x})$ is the frequency modulation (FM) function of $s(\mathbf{x})$. The model (1) is ill-posed in the sense that, for any given $s(\mathbf{x})$, there are uncountably infinitely many distinct pairs of functions $a(\mathbf{x})$, $\varphi(\mathbf{x})$ that satisfy (1). We are most interested in pairs such that both the AM and the FM function are locally smooth, the local image contrast is manifest in $a(\mathbf{x})$, and the local texture orientation and pattern spacing are manifest in $\nabla\varphi(\mathbf{x})$.

In contrast to the real case, for any complex-valued image $z(\mathbf{x}): \mathbb{R}^2 \rightarrow \mathbb{C}$, we may write

$$z(\mathbf{x}) = s(\mathbf{x}) + jq(\mathbf{x}) = a(\mathbf{x}) \exp[j\varphi(\mathbf{x})], \quad (2)$$

$$q(\mathbf{x}) = a(\mathbf{x}) \sin[\varphi(\mathbf{x})]. \quad (3)$$

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. Author E-mail: nick.mould@gmail.com, chuong@ou.edu, Colin.M.Johnston@noaa.gov, joebob@ou.edu.

Unlike (1), the modulating functions $a(\mathbf{x})$ and $\nabla\varphi(\mathbf{x})$ in (2) are unique up to sets of (Lebesgue) measure zero. Indeed, one may verify by direct calculation that, for $z(\mathbf{x})$ in (2), we have

$$\nabla\varphi(\mathbf{x}) = \operatorname{Re} \left[\frac{\nabla z(\mathbf{x})}{jz(\mathbf{x})} \right], \quad (4)$$

$$a(\mathbf{x}) = |z(\mathbf{x})|. \quad (5)$$

Intuitively, the FM algorithm (4) “works” because the function e^u is a fixed point (*e.g.*, eigenfunction) of the derivative operator. Thus, $de^u = e^u du$, from which (4) follows immediately. It is also worth noting that the imaginary component of the quantity $\nabla z(\mathbf{x})/jz(\mathbf{x})$ is $\nabla a(\mathbf{x})/a(\mathbf{x})$,⁴ which is a multidimensional extension of the instantaneous bandwidth as defined by Cohen.⁵

Because (4) and (5) provide unique solutions for the AM and FM functions of any complex-valued image, we consider the problem of associating a unique pair of modulating functions to the real-valued image $s(\mathbf{x})$ in (1) as equivalent to the problem of constructing the complex extension $z(\mathbf{x})$ given by (2); *e.g.*, disambiguation of the real-valued demodulation problem is equivalent to selecting an imaginary part $q(\mathbf{x})$ to associate with $s(\mathbf{x})$. How one should best select $q(\mathbf{x})$ depends on the application and the constraints at hand. In the absence of competing constraints, we have argued^{1,2,6,7} that there are strong physical motivations for choosing $q(\mathbf{x})$ to be the partial Hilbert transform of $s(\mathbf{x})$ according to

$$q(\mathbf{x}) = \mathcal{H}[s(\mathbf{x})] = \frac{1}{\pi x_1} * s(\mathbf{x}) = \frac{1}{\pi} \operatorname{pv} \int_{\mathbb{R}} s(\mathbf{x} - [\xi \ 0]^T) \frac{d\xi}{\xi}, \quad (6)$$

where the asterisk indicates 1-D convolution with respect to x_1 and the integral is interpreted as a Cauchy principal value. With this approach, the complex image $z(\mathbf{x})$ admits many but not all of the attractive properties of the 1-D analytic signal.^{8,9} In particular, it satisfies the following multidimensional extensions of the properties proposed by Vakman¹⁰ as desirable for the 1-D amplitude, phase, and frequency.

1. *Continuity*: if $\hat{s}(\mathbf{x}) = s(\mathbf{x}) + \delta s$, where δs is a small variation, then the modulating functions of $\hat{s}(\mathbf{x})$ are the same as those of $s(\mathbf{x})$ in the limit as $\|\delta s\| \rightarrow 0$.
2. *Amplitude Homogeneity*: if $a(\mathbf{x})$ and $\nabla\varphi(\mathbf{x})$ are the modulating functions of $s(\mathbf{x})$ and if $c > 0$ is a positive real constant, then the modulating functions of the image $cs(\mathbf{x})$ are given by $ca(\mathbf{x})$ and $\nabla\varphi(\mathbf{x})$.
3. *Harmonic Correspondence*: if $\boldsymbol{\Omega} \in \mathbb{R}^2$ and $s(\mathbf{x}) = \cos(\boldsymbol{\Omega}^T \mathbf{x})$, then $a(\mathbf{x}) = 1$ and $\nabla\varphi(\mathbf{x}) = \boldsymbol{\Omega}$; *e.g.*, $q(\mathbf{x}) = \sin(\boldsymbol{\Omega}^T \mathbf{x})$.

Moreover, with $q(\mathbf{x})$ defined as in (6) we have that the first moment of the multidimensional instantaneous frequency $\nabla\varphi(\mathbf{x})$ in (4) with respect to the distribution $|z(\mathbf{x})|^2/\|z(\mathbf{x})\|_{L^2}^2$ of signal energy in space precisely coincides with the first moment of the Fourier frequency with respect to the distribution $|Z(\boldsymbol{\Omega})|^2/\|Z(\boldsymbol{\Omega})\|_{L^2}^2$ of signal energy in the Fourier plane:

$$\int_{\mathbb{R}^2} \nabla\varphi(\mathbf{x}) |z(\mathbf{x})|^2 / \|z(\mathbf{x})\|_{L^2}^2 \, d\mathbf{x} = (2\pi)^{-2} \int_{\mathbb{R}^2} \boldsymbol{\Omega} |Z(\boldsymbol{\Omega})|^2 / \|Z(\boldsymbol{\Omega})\|_{L^2}^2 \, d\boldsymbol{\Omega}. \quad (7)$$

A mathematically consistent single component model of the form (1) can always be computed for any given real-valued $s(\mathbf{x})$. However, for most images of practical interest, the single component model (1) will not admit locally smooth modulating functions that correspond well to visual perception of the image. Therefore, when the goal is to use the computed modulations for image analysis or modulation domain signal processing, it is generally necessary to consider the image instead to be a sum of several components of the form (1), wherein each component has locally smooth AM and FM functions. Thus, we have the real-valued multi-component model

$$s(\mathbf{x}) = \sum_{p=1}^P s_p(\mathbf{x}) = \sum_{p=1}^P a_p(\mathbf{x}) \cos[\varphi_p(\mathbf{x})], \quad (8)$$

where $s_p(\mathbf{x}) \triangleq a_p(\mathbf{x}) \cos[\varphi_p(\mathbf{x})]$. Since the partial Hilbert transform (6) is linear, it may be applied directly to $s(\mathbf{x})$ to generate imaginary components at once for all P components $s_p(\mathbf{x})$ in (8) according to

$$\begin{aligned} z(\mathbf{x}) &= s(\mathbf{x}) + j\mathcal{H}[s(\mathbf{x})] = \sum_{p=1}^P z_p(\mathbf{x}) = \sum_{p=1}^P a_p(\mathbf{x}) \exp[j\varphi_p(\mathbf{x})] \\ &= \sum_{p=1}^P s_p(\mathbf{x}) + j\mathcal{H}[s_p(\mathbf{x})], \end{aligned} \quad (9)$$

where $z_p(\mathbf{x}) \triangleq s_p(\mathbf{x}) + jq_p(\mathbf{x})$, $q_p(\mathbf{x}) = \mathcal{H}[s_p(\mathbf{x})]$, and $\mathcal{H}[s_p(\mathbf{x})]$ is defined by (6). The multi-component models (8) and (9) provide great flexibility in the sense that arbitrarily complicated images $s(\mathbf{x})$ can be modeled as superpositions of *locally coherent* AM-FM functions. However, in practice they also introduce an additional level of ill-posedness to the AM-FM modeling problem, since there is no unique solution for how the image $s(\mathbf{x})$ should be decomposed into components $s_p(\mathbf{x})$.

In (9), if we take all the $a_p(\mathbf{x})$ constant and all the $\varphi_p(\mathbf{x})$ linear, then the FM functions $\nabla\varphi_p(\mathbf{x})$ are all constant and the AM-FM components $z_p(\mathbf{x})$ coincide with stationary Fourier components of $z(\mathbf{x})$. Hence if $\mathbf{m} \in \mathbb{Z}^2$ and $s(\mathbf{m})$ is a discrete image of finite extent, then at least one discrete multi-component AM-FM representation of the form (9) exists, since the DFT is in fact a multicomponent AM-FM model where all the modulating functions are constant. Generally, other multi-component AM-FM models will also exist that are more desirable in the sense of having fewer components and of admitting AM and FM functions that better correspond to visual perception of the salient nonstationary image structures. Two main approaches for obtaining the components $s_p(\mathbf{x})$ from the image $s(\mathbf{x})$ have been investigated. First, one can apply some kind of regression or matching pursuits type of algorithm to iteratively extract components $s_p(\mathbf{x})$.¹¹⁻¹⁵ The empirical mode decomposition is an example of this type of approach. The obtained components can then be demodulated using (4) and (5). While techniques of the first type show great promise, we will not consider them further in this paper because they are generally less mature than techniques belonging to the second main approach.

In the second main approach, one applies a multiband filterbank to obtain the components $s_p(\mathbf{x})$. The filterbank channels need to be well localized spatially to faithfully capture the local nonstationary image structure, but also spectrally well localized to resolve multiple image components that are closely spaced in frequency. For this reason, the use of Gabor filters, which in the continuous domain case uniquely realize the Heisenberg-Weyl inequality lower limit on 2-D joint resolution,¹⁶ has been popular.^{2,3,17-20} Although the Gabor filters admit many attractive properties for AM-FM image analysis,¹ perfect reconstruction is not among them. Indeed, it is generally true that, for any filter family, desirable properties such as good joint localization and local coherency of the impulse and frequency responses are goals that conflict with that of perfect reconstruction. Recently, we have constructed Gabor-like perfect reconstruction filterbanks and used them to perform signal processing in the modulation domain.^{21,22} Doing this invariably requires sacrificing desirable properties of the filterbank in order to obtain perfect reconstruction, however. Like many classical problems in computer vision, target tracking is an “analysis only” application: image analysis is required but image synthesis is not. Thus, a perfect reconstruction system is not required for the target tracking application and we restrict our attention here to multiband Gabor filterbanks that provide many desirable properties^{1,23} but not perfect reconstruction.

Let $g_p(\mathbf{x})$ and $G_p(\boldsymbol{\Omega})$ be the impulse and frequency responses of filterbank channel p . Then the channel response is given by

$$y_p(\mathbf{x}) = z(\mathbf{x}) * g_p(\mathbf{x}) \approx z_p(\mathbf{x}) * g_p(\mathbf{x}) = \int_{\mathbb{R}^2} a_p(\mathbf{u}) \exp[j\varphi(\mathbf{u})] g_p(\mathbf{x} - \mathbf{u}) d\mathbf{u}. \quad (10)$$

Unfortunately, (10) does not lead to a closed form solution for the AM function a_p and the FM function $\nabla\varphi_p$. Therefore, it becomes necessary to *estimate* the modulating functions a_p and $\nabla\varphi_p$ of component s_p from the channel response y_p , which is in general different from s_p in that a shift and scale has been incurred in passing the component through the filterbank channel. We have shown that the modulating functions of component s_p

can be reasonably estimated with bounded errors using the approximate demodulation algorithm²

$$\nabla\varphi_p(\mathbf{x}) = \operatorname{Re} \left[\frac{\nabla y_p(\mathbf{x})}{j y_p(\mathbf{x})} \right], \quad (11)$$

$$a_p(\mathbf{x}) = \left| \frac{y_p(\mathbf{x})}{G_p[\nabla\varphi_p(\mathbf{x})]} \right|. \quad (12)$$

We have also developed discrete versions of (11) and (12) that are applicable to digital images.^{1,2}

3. DUAL DOMAIN DETECTION AND ONLINE TRACK CONSISTENCY CHECKS

We consider an imaging infrared detector that delivers a sequence of digital video frames $s^k(\mathbf{m})$, where $k \in \mathbb{N}$ is the temporal index and $\mathbf{m} = [m_1 \ m_2]^T \in \mathbb{Z}^2$. We assume that the target location and magnification in the initial frame are known. This information could come from an independent detection process such as matched filtering²⁴ against a library of *a priori* known signatures or from a manual designation of a previously unknown target type by a “human in the loop.” As a baseline, we consider a normalized correlation tracker operating on the raw video frames; we refer to this as the *pixel domain* tracker to differentiate it from the modulation domain tracker to be introduced later in this section. In the baseline pixel domain system, tracking is accomplished by computing the normalized correlation function $\gamma_{\text{pix}}(\mathbf{m}, k)$ between the current target template and the currently observed frame $s^k(\mathbf{m})$.²⁵ A detection is declared at the pixel coordinates of the global peak in the normalized correlation function, *e.g.*, $\mathcal{C}_{\text{pix}}(k) = \operatorname{argmax}_{\mathbf{m}} \gamma_{\text{pix}}(\mathbf{m}, k)$. If the noise and clutter are not too severe, then tracking can be accomplished by simply taking the detection $\mathcal{C}_{\text{pix}}(k)$ as the track centroid. In practical combat scenarios, however, the target to clutter ratio is likely to be poor at least at times and so it is usually preferable to model the peak of the correlation function as a noisy measurement to be used as input to a Kalman or particle filter.^{26–29} Employing a more sophisticated track filter in this way provides an additional advantage in that the required computational load can be reduced by restricting the calculation of the normalized correlation function to a neighborhood about the predicted track centroid.

In most scenarios of practical interest, the target signature will exhibit nonstationary evolution with time, including magnification, rotation, and pose. As the observed target signature evolves, the quality of the correlation based detection process degrades in the sense that the template eventually becomes stale and fails to correlate well with the actual target; when this occurs, the track may become locked onto clutter features or may simply fail to detect the correct target location. In either case, track loss typically results. Thus, it is generally necessary to devise some means for identifying that the detection process has been compromised and for updating the template when this occurs.^{30–33} Two straightforward strategies are to update the template every frame and to update every K frames, where $K \in \mathbb{Z}$. However, neither one of these straightforward strategies will provide satisfactory performance against difficult IR sequences such as the well-known AMCOM closure sequences.^{26,34–36} If the template is updated every frame, it will invariably overadapt with the result that the tracker will lock onto stationary clutter features. Relatively improved performance can be obtained with the fixed interval update strategy, but some means must be devised for determining the interval K . With *a priori* knowledge, one can choose a K that will result in successful tracking for some AMCOM sequences. However, we know of no systematic way for determining K .

Our main goal in this paper is to introduce a new dual domain detection process that will facilitate identification of a stale template. When the dual domain detector indicates deterioration of the pixel domain detection process, the target template should be updated. This approach is quite general and could be combined with a variety of different tracking algorithms. Hence, our attention here is focused on dual domain detection as a means of track consistency checking rather than on any particular tracking algorithm. We analyze each video frame $s^k(\mathbf{m})$ with an 18-channel Gabor filterbank which is depicted in the 2-D frequency plane in Fig. 1(a). In practice, we combine the linear translation invariant transform (6) in series with each filterbank channel so that the complex components (9) and the channel responses (10) can be generated simultaneously with $P = 18$ pointwise spectral multiplications. Discrete versions of the demodulation algorithms (11) and (12) are applied to all 18 channel responses $y_p(\mathbf{m})$ to estimate the modulating functions $a_p^k(\mathbf{x})$ and $\nabla\varphi_p^k(\mathbf{m})$, where superscript k denotes the time index. We then convert the FM vector fields to polar coordinates by computing for each

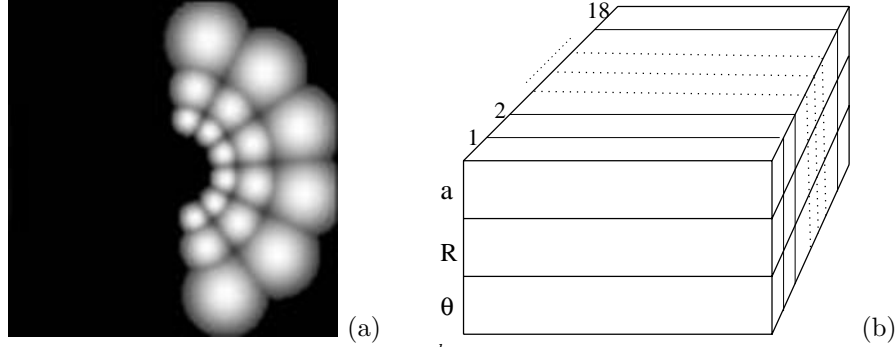


Figure 1. AM-FM image model for a digital video frame $s^k(\mathbf{m})$. (a) Frequency response of 18-channel Gabor filterbank. (b) Modulation domain image planes for the AM and polar FM vectors of all 18 channels.

component the magnitude frequency $r_p^k(\mathbf{m}) = |\nabla\varphi_p^k(\mathbf{m})|$ and orientation $\theta_p^k(\mathbf{m}) = \arg \nabla\varphi_p^k(\mathbf{m})$. This results in a modulation domain image representation comprising 54 image planes $a_p^k(\mathbf{m})$, $r_p^k(\mathbf{m})$, and $\theta_p^k(\mathbf{m})$ for $1 \leq p \leq 18$ as depicted in Fig. 1(b).

The spatial support of the initial pixel domain target template designated in frame $s^1(\mathbf{m})$ is used to simultaneously extract 54 modulation domain templates from $a_p^1(\mathbf{m})$, $r_p^1(\mathbf{m})$, and $\theta_p^1(\mathbf{m})$. For each frame $s^k(\mathbf{m})$, normalized correlation functions are computed in all 54 modulation domain image planes depicted in Fig. 1(b). These 54 normalized correlation functions are then averaged to construct a composite correlation function $\gamma_{\text{mod}}(\mathbf{m}, k)$ which we refer to as the *modulation domain correlation function*. The modulation domain target detection for frame k is declared at the peak of the modulation domain correlation function according to $\mathcal{C}_{\text{mod}}(k) = \text{argmax}_{\mathbf{m}} \gamma_{\text{mod}}(\mathbf{m}, k)$. Two comments are in order concerning $\mathcal{C}_{\text{mod}}(k)$ and $\gamma_{\text{mod}}(\mathbf{m}, k)$. First, one could alternatively calculate $\mathcal{C}_{\text{mod}}(k)$ by averaging the pixel coordinates of the 54 peaks of the individual modulation domain normalized correlation functions computed from $a_p^k(\mathbf{m})$, $r_p^k(\mathbf{m})$, and $\theta_p^k(\mathbf{m})$. We have thoroughly tested this approach against the AM-COM sequences. The empirical conclusion is that better agreement with the ground truth target centroids is obtained by averaging the 54 correlation functions first and then computing the maximum $\mathcal{C}_{\text{mod}}(k)$ as opposed to averaging the 54 individual maxima. Second, one could devise consistency checks or quality metrics for the 54 modulation domain image planes and use them to perform an unequally weighted average of the 54 individual correlation functions. That is an interesting and promising idea that we have yet to investigate.

3.1 Online Track Consistency Checks

The AM-FM measurement shown in Fig. 1(b) provides a transform domain representation of the frame $s^k(\mathbf{m})$ that is distinctly different from the pixel domain representation. In particular, the AM-FM functions provide a rich description of the local nonstationary texture structure. We have argued previously that targets tend to exhibit a higher degree of organization or coherency in the modulation domain as compared to backgrounds and clutter and also that target-background class separability is often enhanced in the modulation domain as compared to the pixel domain.^{37,38} In theory, one could perform tracking using the pixel domain detection $\mathcal{C}_{\text{pix}}(k)$ or the modulation domain detection $\mathcal{C}_{\text{mod}}(k)$. Through exhaustive empirical testing against 36 AMCOM sequences for which ground truth is available we have determined that $\mathcal{C}_{\text{pix}}(k)$ and $\mathcal{C}_{\text{mod}}(k)$ are both approximately unbiased detectors. However, the error variance (*e.g.*, spatial imprecision) of $\mathcal{C}_{\text{mod}}(k)$ is greater due to the inherent spreading, *i.e.*, spatial averaging, that occurs as a result of applying the Gabor filters in computing the AM-FM image model. Based on this empirical evidence, in practice we use only $\mathcal{C}_{\text{pix}}(k)$ and not $\mathcal{C}_{\text{mod}}(k)$ as input to a higher level track processor.

However, because the pixel domain and modulation domain representations provide distinctly different descriptions of the true 3-D scene, the pixel domain and modulation domain correlation-based detection processes tend to fail in substantially different ways. This observation suggests that by quantifying the *difference* in the failure modes of $\mathcal{C}_{\text{pix}}(k)$ and $\mathcal{C}_{\text{mod}}(k)$ one can obtain an indication of when the pixel domain detector has deteriorated due to a stale template. Again through an exhaustive study of the AMCOM sequences, we observed many cases where nonstationary evolution of the target signature causes a failure in $\mathcal{C}_{\text{pix}}(k)$ but not in $\mathcal{C}_{\text{mod}}(k)$,

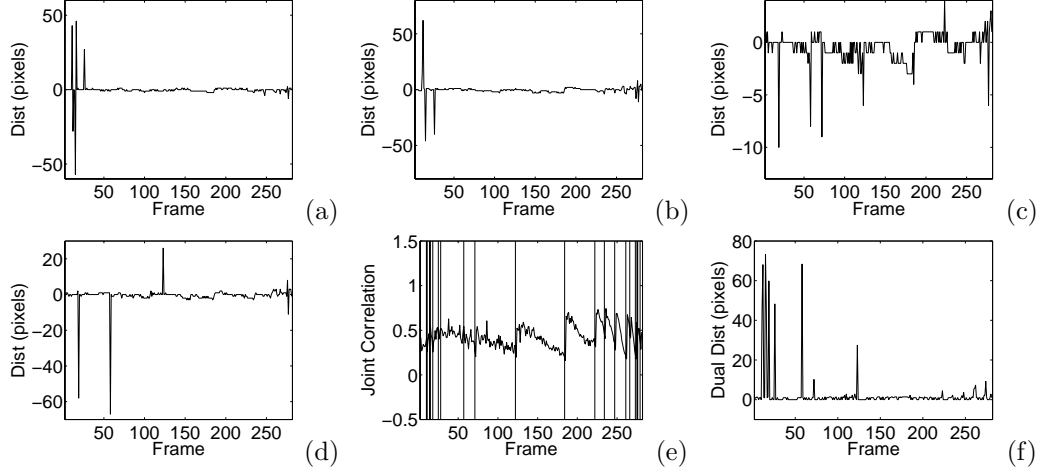


Figure 2. Detection results for *rng14_15*. (a),(b) Horizontal and vertical errors in $\mathcal{C}_{\text{pix}}(k)$ (pixels). (c),(d) Horizontal and vertical errors in $\mathcal{C}_{\text{mod}}(k)$ (pixels). (e) $\gamma_{\text{joint}}(\mathbf{m}, k)$; vertical lines indicate template updates. (f) $\delta_{\text{dual}}(k)$.

a failure in $\mathcal{C}_{\text{mod}}(k)$ but not in $\mathcal{C}_{\text{pix}}(k)$, and also many cases where both detection processes fail simultaneously. However, it is *exceedingly rare* that the two detectors fail at the same time with *similar failure modes*. Thus, in the difficult AMCOM closure sequences, a stale template condition can be detected about 95% of the time by online consistency checks based on the divergence of $\mathcal{C}_{\text{pix}}(k)$ and $\mathcal{C}_{\text{mod}}(k)$. It should be noted in this regard that a track quality metric based solely on agreement between the observed target location and the predicted location delivered by a high level track filter does not provide a sufficient indication of when the pixel domain detection process has been compromised due to a stale template. Indeed, the stale template will often correlate well with elements of the clutter and background; when this occurs, the track gradually locks onto the clutter as opposed to the target and this can occur without any substantial disagreement between the predicted and detected target locations. This occurs because it is the detection process that has been compromised. Clearly, this problem is exacerbated in cases where the template is dynamic and is permitted to overadapt.

Here, we propose two new online track consistency checks capable of revealing a compromised detection process. For the first check, we quantify the quality of the detection by the quantity $\gamma_{\text{joint}}(\mathbf{m}, k) = |\gamma_{\text{pix}}(\mathbf{m}, k)\gamma_{\text{mod}}(\mathbf{m}, k)|$, which we refer to as the *joint correlation product*. As one or both of $\mathcal{C}_{\text{pix}}(k)$ and $\mathcal{C}_{\text{mod}}(k)$ begin to fail, the peak value of $\gamma_{\text{joint}}(\mathbf{m}, k)$ is reduced. Thus, divergence of the two detection processes can be revealed by applying a threshold to the maximum of $\gamma_{\text{joint}}(\mathbf{m}, k)$. For the second track consistency check, the quality of the detection process is quantified by the Euclidean distance $\delta_{\text{dual}}(k) = \|\mathcal{C}_{\text{pix}}(k) - \mathcal{C}_{\text{mod}}(k)\|_{\ell^2}$, here referred to as the *dual domain distance*. For the AMCOM sequences, we generally consider that a stale template condition exists if $\delta_{\text{dual}}(k) > 4$ pixels.

3.2 Examples

We performed experiments to study the degradation of the pixel domain and modulation domain detection processes due to nonstationary target signature evolution for the 36 AMCOM closure sequences shown in the first column of Table 1. For each sequence, we took the target location and template size from ground truth in the first frame. We ran the dual domain detection process until a stale template condition was declared due to an error of more than four pixels with respect to ground truth in one or both of $\mathcal{C}_{\text{pix}}(k)$ and $\mathcal{C}_{\text{mod}}(k)$. Statistics were then recorded to characterize the failure modes of $\mathcal{C}_{\text{pix}}(k)$ and $\mathcal{C}_{\text{mod}}(k)$ as well as the effectiveness of the two online consistency checks for detecting the stale template condition. The dual domain detection process was then restarted using ground truth in the frame where the stale template was declared.

For each AMCOM sequence, the third column of Table 1 gives the number of times that the error in $\mathcal{C}_{\text{pix}}(k)$ with respect to ground truth exceeded four pixels but the error in $\mathcal{C}_{\text{mod}}(k)$ did not. Similarly, column four gives the number of times that the error in $\mathcal{C}_{\text{mod}}(k)$ exceeded four pixels but the error in $\mathcal{C}_{\text{pix}}(k)$ did not. The number of times that the detection process failed simultaneously in both domains but with different failure modes is

Table 1. Divergence statistics for pixel and modulation domain detectors for 36 AMCOM sequences.

Sequence Name	Number of Frames	Divergence Domain		Both Domains		Effectiveness %
		Pixel	Modulation	Asimilar	Similar	
rng14_15	281	10	6	0	3	84
rng15_20	171	14	5	0	0	100
rng15_NS	320	12	10	0	6	79
rng16_04	208	5	3	0	2	80
rng16_07	76	3	2	0	1	83
rng16_08	290	4	3	0	1	88
rng16_18	291	24	8	2	8	81
rng17_01	371	1	2	0	0	100
rng17_02	301	12	14	2	0	100
rng17_20	35	4	1	0	0	100
rng18_03	447	57	4	1	5	93
rng18_05	213	23	11	6	9	82
rng18_07	260	42	7	3	0	100
rng18_12	300	75	3	5	3	97
rng18_13	238	41	5	1	0	100
rng18_16	271	16	18	5	2	96
rng18_18	207	26	9	3	2	95
rng19_01	240	20	0	0	0	100
rng19_02	270	29	2	4	0	100
rng19_04	270	20	5	0	2	96
rng19_10	265	1	7	0	0	100
rng19_11	58	1	2	0	0	100
rng19_13	368	9	9	3	0	100
rng19_15	348	7	3	1	0	88
rng19_18	193	4	8	5	0	100
rng19_NS	448	0	6	0	0	100
rng20_04	141	57	12	8	0	100
rng20_08	334	21	20	8	7	100
rng20_17	475	37	6	13	0	100
rng20_18	360	51	14	5	0	100
rng20_20	348	3	2	0	0	96
rng21_04	380	11	3	1	0	100
rng21_15	114	39	1	0	0	100
rng21_17	165	22	12	1	0	100
rng22_06	142	5	77	3	4	100
rng22_08	101	4	11	5	0	100

given in the fifth column, while the number of times that failure occurred in both domains with similar failure modes is given in the sixth column. The detection process failures in columns three, four, and five can all be identified in real time by thresholding $\delta_{\text{dual}}(k)$ at four pixels. Only the failures shown in column six of the table are undetectable by the online consistency checks proposed in this paper. This is reflected in the rightmost column of Table 1, which gives the percentage of stale template conditions in each sequence that can be detected by the test $\delta_{\text{dual}}(k) > 4$.

Detailed data for five sequences that we consider among the most difficult in the AMCOM data set are given in Figs. 2-6. In each figure, the (a) and (b) parts give the horizontal and vertical errors in the pixel domain detection $\mathcal{C}_{\text{pix}}(k)$ with respect to ground truth, while the horizontal and vertical errors in the modulation domain detection $\mathcal{C}_{\text{mod}}(k)$ with respect to ground truth are given in the (c) and (d) parts. The detection quality metric $\gamma_{\text{joint}}(\mathbf{m}, k)$ is plotted as a function of k in the (e) part of each figure, where vertical lines indicate the frames where a stale template condition was declared and the detectors were reinitialized from ground truth. Although the vertical axis drops below zero on the (f) part of Fig. 2 in order to show the vertical bars clearly, it should be noted that the $\gamma_{\text{joint}}(\mathbf{m}, k)$ data are strictly non-negative in all five figures. After each template reinitialization, $\gamma_{\text{joint}}(\mathbf{m}, k)$ shows a clear decreasing trend due to degradation of the detection process because of nonstationary

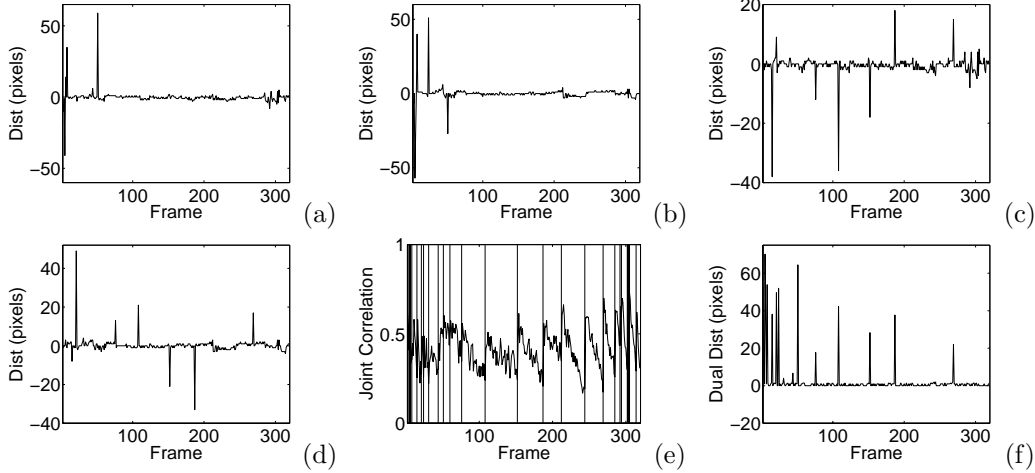


Figure 3. Detection results for *rng15_NS*. (a),(b) Horizontal and vertical errors in $\mathcal{C}_{\text{pix}}(k)$ (pixels). (c),(d) Horizontal and vertical errors in $\mathcal{C}_{\text{mod}}(k)$ (pixels). (e) $\gamma_{\text{joint}}(\mathbf{m}, k)$; vertical lines indicate template updates. (f) $\delta_{\text{dual}}(k)$.

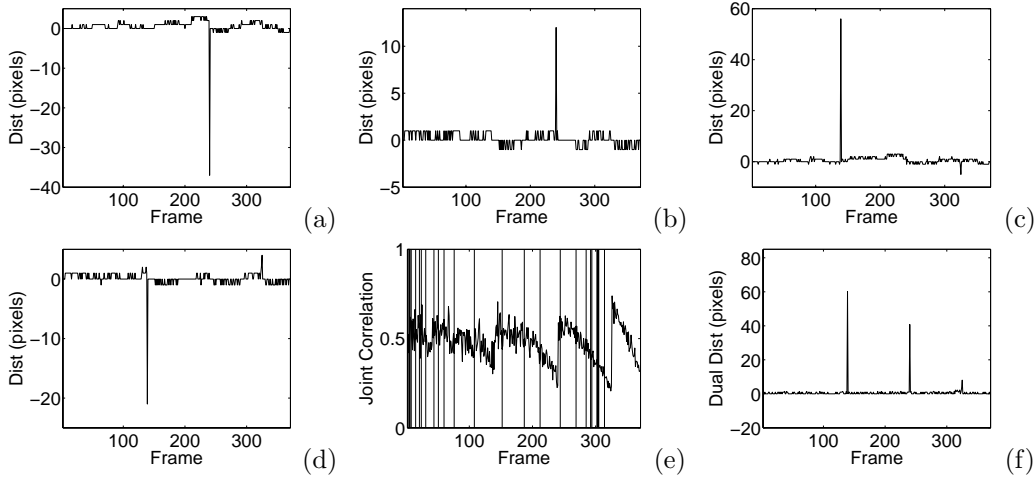


Figure 4. Detection results for *rng17_01*. (a),(b) Horizontal and vertical errors in $\mathcal{C}_{\text{pix}}(k)$ (pixels). (c),(d) Horizontal and vertical errors in $\mathcal{C}_{\text{mod}}(k)$ (pixels). (e) $\gamma_{\text{joint}}(\mathbf{m}, k)$; vertical lines indicate template updates. (f) $\delta_{\text{dual}}(k)$.

target signature evolution, as expected. Finally, the metric $\delta_{\text{dual}}(k)$ is given in the (f) part of Figs. 2-6. Thus, the vertical lines in the (e) part of each figure occur where the quantity in the (f) part exceeds four pixels.

Examples of interesting frames where failures occurred in one or both domains are shown in Figs. 7-9, again for sequences that we consider among the most difficult. In each figure, the black rectangle shows the size of the pixel domain template and is centered at $\mathcal{C}_{\text{pix}}(k)$, while the white rectangle shows the size of the modulation domain template and is centered at $\mathcal{C}_{\text{mod}}(k)$. Frames from the sequence *rng14_15* appear in Fig. 7. Fig. 7(a) shows frame $s^{10}(\mathbf{m})$, where a stale template condition was declared due to an error of greater than four pixels with respect to ground truth in the pixel domain detector but not in the AM-FM detector. Fig. 7(b) shows frame 184, where failure occurred in the modulation domain but not in the pixel domain. Finally, Fig. 7(c) shows frame 279, where both detectors failed simultaneously. Interestingly, the failure mode was the same in both domains, so that $\delta_{\text{dual}}(279) < 4$ pixels. Thus, Fig. 7(c) is an example of a stale template condition that cannot be detected with the online consistency checks proposed in this paper. However, as may be seen from the figure, this failure is not militarily severe, in the sense that the probability of kill is still high with almost any conceivable precision guided munitions system. Similar examples are given for the sequence *rng19_13* in Fig. 8 and for the sequence *rng18_05* in Fig. 9.

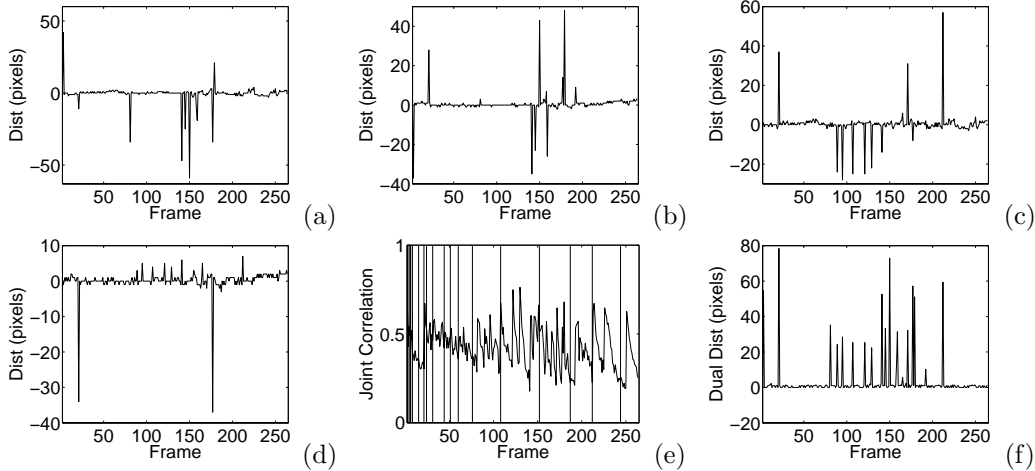


Figure 5. Detection results for *rng19_13*. (a),(b) Horizontal and vertical errors in $\mathcal{C}_{\text{pix}}(k)$ (pixels). (c),(d) Horizontal and vertical errors in $\mathcal{C}_{\text{mod}}(k)$ (pixels). (e) $\gamma_{\text{joint}}(\mathbf{m}, k)$; vertical lines indicate template updates. (f) $\delta_{\text{dual}}(k)$.

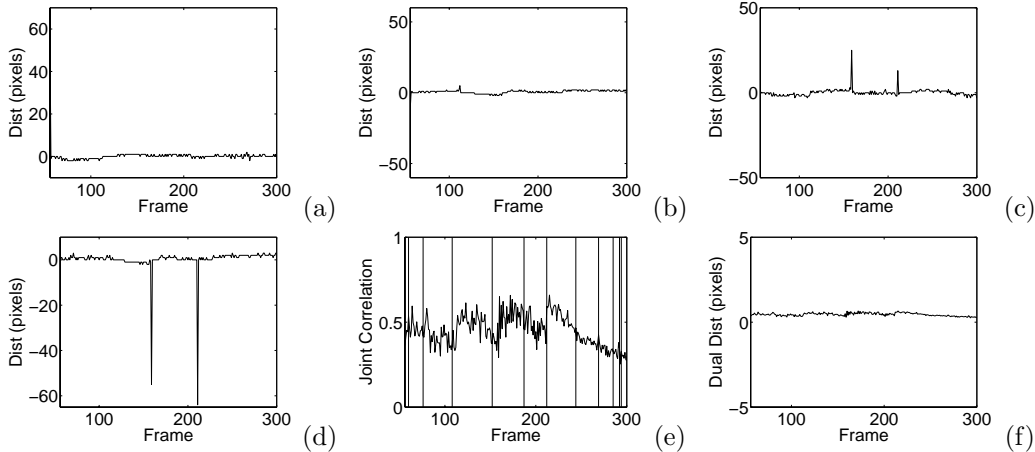


Figure 6. Detection results for *rng21_17*. (a),(b) Horizontal and vertical errors in $\mathcal{C}_{\text{pix}}(k)$ (pixels). (c),(d) Horizontal and vertical errors in $\mathcal{C}_{\text{mod}}(k)$ (pixels). (e) $\gamma_{\text{joint}}(\mathbf{m}, k)$; vertical lines indicate template updates. (f) $\delta_{\text{dual}}(k)$.

4. CONCLUSION

We reviewed the fundamentals of AM-FM image modeling and proposed two dual domain online consistency checks for identifying when a stale template condition is compromising the detection process in an infrared target tracking system. These consistency checks are independent of any particular track filter and hence can be applied for track quality monitoring with a variety of different tracking systems. To demonstrate the practical utility of these new consistency checks against a realistic test case, we ran a pixel domain normalized correlation tracker against the AMCOM sequence *rng19_13*. We consider this sequence to be one of the most difficult in the data set because there are multiple targets present in the frame, each target exhibits complicated kinematics, and there is *considerable* nonstationary evolution of the target signatures throughout the sequence. The problem was to track the lead vehicle. As shown by the sample frames in Fig. 10, even this simple track filter is successful in maintaining track lock throughout this extremely difficult sequence when the online track consistency checks proposed in this paper are used to initiate template refresh operations.

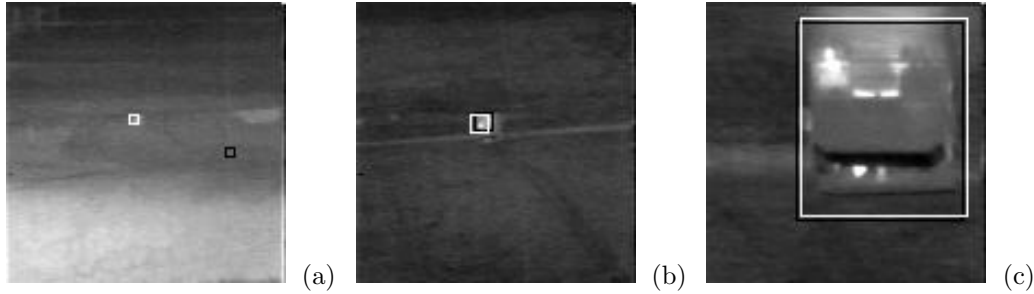


Figure 7. Example detector failures for *rng14-15*. The pixel domain template is shown in black and the AM-FM template is shown in white. (a) Failure of $\mathcal{C}_{\text{pix}}(k)$ but not $\mathcal{C}_{\text{mod}}(k)$ at $k = 10$. (b) Failure of $\mathcal{C}_{\text{mod}}(k)$ but not $\mathcal{C}_{\text{pix}}(k)$ at $k = 184$. (c) Failure of both $\mathcal{C}_{\text{pix}}(k)$ and $\mathcal{C}_{\text{mod}}(k)$ at $k = 279$. In this case the failure mode is the same in both domains, so the consistency checks proposed here are incapable of detecting this stale template condition. This failure is not serious from a tactical standpoint, however, since the probability of kill is still high.



Figure 8. Example detector failures for *rng19-13*. (a) Failure of $\mathcal{C}_{\text{pix}}(k)$ but not $\mathcal{C}_{\text{mod}}(k)$ at $k = 3$. (b) Failure of both $\mathcal{C}_{\text{pix}}(k)$ and $\mathcal{C}_{\text{mod}}(k)$ at $k = 21$; the consistency checks still detect the stale template condition, however, since the failure modes are different in the two domains. (c) Failure of $\mathcal{C}_{\text{mod}}(k)$ but not $\mathcal{C}_{\text{pix}}(k)$ at $k = 89$.

REFERENCES

1. J. P. Havlicek, P. C. Tay, and A. C. Bovik, "AM-FM image models: Fundamental techniques and emerging trends," in *Handbook of Image and Video Processing*, A. C. Bovik, ed., pp. 377–395, Elsevier Academic Press, Burlington, MA, 2nd ed., 2005.
2. J. P. Havlicek, D. S. Harding, and A. C. Bovik, "Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models," *IEEE Trans. Image Proc.* **9**, pp. 227–242, Feb. 2000.
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**, pp. 691–712, Mar. 1992.
4. J. P. Havlicek, D. S. Harding, and A. C. Bovik, "Multicomponent multidimensional signals," *Multidimensional Syst. and Signal Proc.* **9**, pp. 391–398, Oct 1998.
5. L. Cohen, *Time-Frequency Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1995.
6. J. P. Havlicek, J. W. Havlicek, and A. C. Bovik, "The analytic image," in *Proc. IEEE Int'l. Conf. Image Proc.*, **2**, pp. 446–449, (Santa Barbara, CA), Oct. 26-29, 1997.
7. J. P. Havlicek, J. W. Havlicek, N. D. Mamuya, and A. C. Bovik, "Skewed 2D Hilbert transforms and computed AM-FM models," in *Proc. IEEE Int'l. Conf. Image Proc.*, pp. 602–606, (Chicago, IL), Oct. 4-7, 1998.
8. D. Gabor, "Theory of communication," *J. Inst. Elect. Eng. London* **93**(III), pp. 429–457, 1946.
9. J. Ville, "Théorie et applications de la notation de signal analytique," *Cables et Transmission* **2A**, pp. 61–74, 1948. translated from the French in I. Selin, "Theory and applications of the notion of complex signal," Tech. Rept. T-92, The RAND Corporation, Santa Monica, CA, Aug. 1958.
10. D. Vakman, "On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency," *IEEE Trans. Signal Proc.* **44**, pp. 791–797, Apr. 1996.

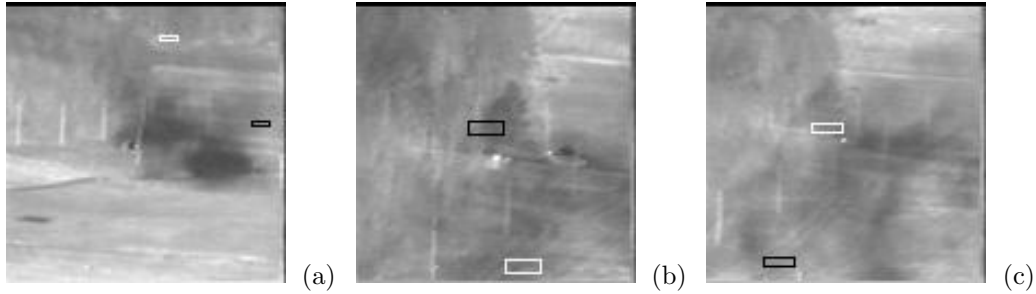


Figure 9. Example detector failures for *rng18_05*. (a) Failure of both $\mathcal{C}_{\text{pix}}(k)$ and $\mathcal{C}_{\text{mod}}(k)$ at $k = 525$. The consistency checks detect the failures in this case because the failure modes are different in the two domains. (b) Failure of $\mathcal{C}_{\text{mod}}(k)$ but not $\mathcal{C}_{\text{pix}}(k)$ at $k = 698$. (c) Failure of $\mathcal{C}_{\text{pix}}(k)$ but not $\mathcal{C}_{\text{mod}}(k)$ at $k = 722$.



Figure 10. Six tracked frames from the sequence *rng19_13*. Ground truth information is not used after the initial target designation in the first frame. The template is refreshed when $\delta_{\text{dual}}(k) > 4$. With this template update strategy, even a simple pixel domain normalized correlation tracker successfully tracks the lead vehicle throughout the entirety of this extremely difficult sequence.

11. N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. Lond. A* **454**, pp. 903–995, 1998.
12. Z. Liu, H. Wang, and L. Peng, "Texture segmentation using directional empirical mode decomposition," in *Proc. IEEE Int'l. Conf. Image Proc.*, **1**, pp. 279–281, (Singapore), Oct. 24–27, 2004.
13. S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Proc.* **41**, pp. 3397–3415, Dec. 1993.
14. R. Neff and A. Zakhori, "Matching pursuit video coding – part I: Dictionary approximation," *IEEE Trans. Circuits Syst. Video Technol.* **12**, pp. 13–26, Jan. 2002.
15. B. Krishnapuram, J. Sichina, and L. Carin, "Physics-based detection of targets in SAR imagery using support vector machines," *IEEE Sensors J.* **3**, pp. 147–157, Dec. 1993.

16. J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Am. A* **2**, pp. 1160–1169, Jul. 1985.
17. P. Maragos and A. C. Bovik, "Image demodulation using multidimensional energy separation," *J. Opt. Soc. Amer. A* **12**, pp. 1867–1876, Sep. 1995.
18. T. Ezzat, J. Bouvrie, and T. Poggio, "AM-FM demodulation of spectrograms using localized 2D max-Gabor analysis," in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Process.*, **IV**, pp. 1061–1064, (Honolulu, HI), Apr. 15-20, 2007.
19. F. Tang and H. Tao, "Non-orthogonal binary expansion of Gabor filters with applications in object tracking," in *Proc. IEEE Workshop Motion, Video Comput.*, p. 24 (6 pages), (Austin, TX), Feb. 23-24 2007.
20. J. K. Kamarainen, V. Kyrki, and H. Kälviäinen, "Invariance properties of Gabor filter-based features — overview and applications," *IEEE Trans. Image Proc.* **15**, pp. 1088–1099, May 2006.
21. R. A. Sivley and J. P. Havlicek, "Perfect reconstruction AM-FM image models," in *Proc. IEEE Int'l. Conf. Image Proc.*, pp. 2125–2128, (Atlanta, GA), Oct. 8-11, 2006.
22. C. T. Nguyen, R. A. Sivley, and J. P. Havlicek, "First results in AM-FM image filtering," in *Proc. IEEE Southwest Symp. Image Anal., Interp.*, (Santa Fe, NM), Mar. 24-27, 2008, *submitted*.
23. J. P. Havlicek, A. C. Bovik, and D. Chen, "AM-FM image modeling and Gabor analysis," in *Visual Information Representation, Communication, and Image Processing*, C. W. Chen and Y. Zhang, eds., pp. 343–385, Marcel Dekker, New York, 1999.
24. G. L. Turin, "An introduction to digital matched filters," *Proc. IEEE* **64**, pp. 1092–1112, Jul. 1976.
25. 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 *Applications of Digital Image Processing XXVIII*, A. G. Tescher, ed., *Proc. SPIE* **5909**, pp. 590907–1 – 590907–12, 2005.
26. A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.* **38**, pp. 1–45, Dec. 2006.
27. 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**, pp. 174–188, Feb. 2002.
28. M. G. S. Bruno, "Sequential importance sampling filtering for target tracking in image sequences," *IEEE Signal Process. Lett.* **10**(8), pp. 246–249, 2003.
29. M. Bruno, "Bayesian methods for multiaspect target tracking in image sequences," *IEEE Trans. Signal Process.* **52**, pp. 1848–1861, Jul. 2004.
30. I. Matthews, T. Ishikawa, and S. Baker, "The template update problem," *IEEE Trans. Pattern Anal., Machine Intel.* **26**, pp. 810–815, Jun. 2004.
31. L. J. Latecki and R. Mieziako, "Object tracking with dynamic template update and occlusion detection," in *Proc. IEEE Int'l. Conf. Pattern Recog.*, pp. 556–560, (Hong Kong, China), Aug. 20-24, 2006, vol. 1.
32. G. D. Hager 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.*, pp. 403–410, (San Francisco, CA), Jun. 18-20 1996.
33. Z. Peng, Q. Zhang, and A. Guan, "Extended target tracking using projection curves and matching pel count," *Optical Eng.* **46**, pp. 066401–1 – 066401–6, Jun. 2007.
34. J. Khan and M. Alam, "Efficient target detection in cluttered flir imagery," in *Optical Pattern Recognition XVI*, D. Casasent and T.-H. Chao, eds., *Proc. SPIE* **5816**, pp. 39–53, 2005.
35. S. Yi and L. Zhang, "A novel multiple tracking system for UAV platforms," in *ISR Systems and Applications III*, D. Henry, ed., *Proc. SPIE* **6209**, 2006.
36. A. Dawoud, M. Alam, A. Bai, and C. Loo, "Decision fusion algorithm for target tracking in infrared imagery," *Opt. Eng.* **44**(2), 2005.
37. J. P. Havlicek, C. T. Nguyen, and M. Yearly, "Modulation domain infrared target models," in *Targets and Backgrounds XII: Characterization and Representation*, W. R. Watkins and D. Clement, eds., *Proc. SPIE* **6239**, pp. 62390D–1 – 62390D–11, 2006.
38. C. T. Nguyen and J. P. Havlicek, "Modulation domain features for discriminating infrared targets and backgrounds," in *Proc. IEEE Int'l. Conf. Image Proc.*, pp. 3245–3248, (Atlanta, GA), Oct. 8-11, 2006.