

# Adaptive Kalman Filtering for Histogram-based Appearance Learning in Infrared Imagery

Vijay Venkataraman, *Member, IEEE*, Guoliang Fan, *Senior Member, IEEE*, Joseph P. Havlicek, *Senior Member, IEEE*, Xin Fan, *Member, IEEE*, Yan Zhai, *Senior Member, IEEE*, and Mark Yeary, *Senior Member, IEEE*

**Abstract**—Targets of interest in video acquired from imaging infrared sensors often exhibit profound appearance variations due to a variety of factors including complex target maneuvers, ego-motion of the sensor platform, background clutter, *etc.*, making it difficult to maintain a reliable detection process and track lock over extended time periods. Two key issues in overcoming this problem are how to represent the target and how to learn its appearance online. In this work, we adopt a recent appearance model that estimates the pixel intensity histograms as well as the distribution of local standard deviations in both the foreground and background regions for robust target representation. Appearance learning is then cast as an adaptive Kalman filtering (AKF) problem where the process and measurement noise variances are both unknown. We formulate this problem using both covariance matching and, for the first time in a visual tracking application, the recent autocovariance least-squares (ALS) method. Although convergence of the ALS algorithm is guaranteed only for the case of globally wide sense stationary (WSS) process and measurement noises, we demonstrate for the first time that the technique can often be applied with great effectiveness under the much weaker assumption of piecewise stationarity. The performance advantages of the ALS method relative to classical covariance matching are illustrated by means of simulated stationary and nonstationary systems. Against real data, our results show that the ALS-based algorithm outperforms covariance matching as well as traditional histogram similarity-based methods, achieving sub-pixel tracking accuracy against the well-known AMCOM closure sequences and the recent SENSIAC ATR dataset.

**Index Terms**— Appearance learning, histogram-based appearance model, infrared tracking, adaptive Kalman filter

## I. INTRODUCTION

We consider the problem of tracking maneuvering ground targets in infrared (IR) imagery acquired from airborne and ground-based platforms, where the targets of interest are often noncooperative. Such targets frequently exhibit complex, unexpected maneuvers that can be both difficult to model and difficult to track given noisy measurements from a passive sensor. In this paper, we will be thinking primarily in terms of a sensor that operates in the 3-5  $\mu\text{m}$  midwave IR (MWIR) or 8-12  $\mu\text{m}$  longwave IR (LWIR) bands, both of which have been used in production IR systems for a long time.

This work was supported in part by the U.S. Army Research Laboratory and U.S. Army Research Office under grants W911NF-04-1-0221 and W911NF-08-1-0293, and the Oklahoma Center for the Advancement of Science and Technology under awards HR09-030 and HR12-030.

V. Venkataraman and G. Fan are with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, USA (email: vijay2181@yahoo.com; guoliang.fan@okstate.edu). X. Fan is with Dalian University of Technology, China (email: xin.fan@ieee.org). J. Havlicek, Y. Zhai and M. Yeary are with the School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019 USA (email: yan.zhai@ieee.org; {joebob,yeary}@ou.edu).

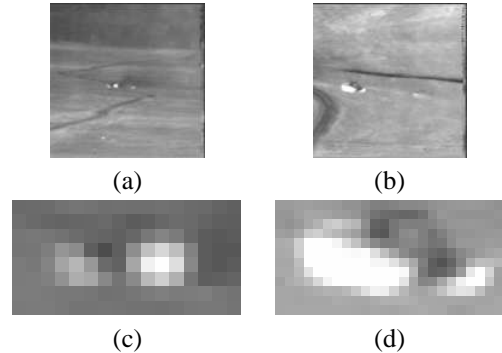


Fig. 1. Nonstationary target signature evolution in AMCOM LWIR run rng18\_17. The lead vehicle is barely visible. The second vehicle is the target of interest. (a) Frame 24. (b) Frame 165. (c)-(d) closeup views of the second target in frames 24 and 165, respectively.

The imagery acquired by such sensors under actual field conditions is typically characterized by strong structured clutter, poor SNR, low target-to-clutter ratios, and strong ego-motion. Particularly for a highly maneuverable target, this implies that the target and background signatures observed at the sensor focal plane array (FPA) may exhibit profound nonstationary variations over relatively short time scales, making it difficult to maintain both a reliable detection process and a robust track lock over longer time scales – phenomena that have been referred to variously as the “drifting problem” in [1], [2], the “template update problem” in [3]–[6], and a “stale template condition” in [7]. These challenges are exemplified by the well-known AMCOM closure sequences<sup>1</sup> [8]–[15] as well as the newly released SENSIAC ATR dataset.<sup>2</sup> One instance of this kind of nonstationary target signature evolution occurs in AMCOM LWIR sequence rng18\_17. Here, an LWIR sensor is situated on an airborne platform that closes on a pair of maneuvering ground vehicles. Frames 24 and 165 are shown in Fig. 1(a) and (b). The target of interest is the second vehicle. A closeup view of this target in frame 24 is given in Fig. 1(c). A closeup view from frame 165 is given in Fig. 1(d). While the second vehicle exhibits a strong signature, the lead vehicle is much dimmer and is in fact barely visible amid the surrounding clutter, demonstrating that brightness alone cannot be used as the sole basis for reliable detection and tracking. Rather, more sophisticated techniques are generally required for representing the target appearance and for adapting to

<sup>1</sup>Available from the Johns Hopkins University Center for Imaging Science (<http://cis.jhu.edu>) and elsewhere.

<sup>2</sup>Available from the Military Sensing Information Analysis Center (<https://www.sensiac.org>).

(e.g., learning) complex appearance changes that occur over time. The choice of a particular target representation often depends on the problem at hand and there exist several learning strategies for each type of representation. In the remainder of this section we consider the general problem and then introduce the specific aspects that will be our focus for the rest of the paper.

#### A. Target representation and appearance learning

Target representations may be broadly categorized as parametric, where a statistical model is typically assumed that captures the key characteristics of the target appearance in a way that facilitates estimation of the model parameters continuously online [16], or non-parametric, where the target appearance is characterized by empirically derived features that can be updated online during tracking [17], [18]. Such features may include kernel-based windows [19]–[21], non-parametric or semiparametric contours [22], templates [20], shape descriptors [19], or local statistics [20], [23] including, e.g., intensity histograms and their moments.

Significant efforts have been directed towards developing methods for online appearance learning [1], [3], [16], [24], [25]. For both parametric and non-parametric approaches, the design of an effective learning strategy is strongly coupled to the choice of features. Drift correction strategies for template tracking were proposed in [3], [26]. A more sophisticated model combining stable, wandering, and outlier components in a Gaussian mixture model (GMM) was proposed in [16], where the model was updated via an expectation maximization (EM) algorithm. GMM-based appearance learning was also applied in [27], where a mean-shift algorithm was used to update the parameters online. These methods rely on elaborate parametric models and are effective for tracking extended targets with large spatial signatures. However, for targets such as those shown in Figs. 1 and 2, there may not be enough pixels on the target to achieve robust and statistically significant parameter estimation.

#### B. Histogram-based appearance learning

Histograms of the pixel intensities have been widely used and were adopted in the appearance models of several recent mean-shift trackers [17], [28]–[30]. Histograms of the local standard deviation (stdev) were also used for mean-shift tracking of IR targets in [23]. The popularity of histogram-based features results at least in part from their simplicity and efficiency, as well as their scale and rotation invariance properties [17], [23], [31]. For histogram-based target representations, appearance learning is generally accomplished by iteratively updating a reference histogram [30], [32], [33]. Typically, the new reference histogram at each iteration is given by a linear weighting of the previous reference histogram and the most recent observation, where the weighting may be based on an appropriate measure of histogram similarity. While such techniques are often effective for adapting the appearance model when the target has a large spatial extent, they can be susceptible to drifting problems, particularly when applied to smaller targets. Alternatively, a method was

proposed in [34] for updating the reference histogram which treats the observed histogram as a realization of a generative model that is a piecewise linear combination of several pairs of histograms computed from representative key appearances of the target. This approach is suitable when the objects to be tracked share substantial similarity (e.g., in certain face and head tracking problems) or when there exists a satisfactory *a priori* means for estimating a meaningful set of key appearance histograms.

An extension of the simple histogram-based appearance learning strategies that has been used to combat the drifting problem involves maintaining explicit appearance models for both the target and the surrounding background. Background information was explicitly incorporated in [35]–[39] to represent the target in terms of features capable of enhancing background discrimination performance. In [14], we proposed a dual foreground-background appearance model comprising four histograms that characterize the pixel intensity distribution and the local distribution of the sample stdev over both the target and the surrounding background. The local stdev statistic amplifies signatures of small and dim targets while minimizing the effect of uniform background clutter. This appearance model will be used by all of the target tracking algorithms considered in this paper. We also found that explicit appearance modeling of background immediately around the target tends to improve the estimation of the target magnification — a problem that is often under treated in part due to the absence of any universal robust and learnable target model [23].

#### C. Adaptive Kalman filtering for histogram learning

As an illustrative example, time traces of the normalized pixel intensity histograms for the target and local background in AMCOM LWIR sequence `rng17_01` are given in Fig. 2 along with several raw video frames. In the early part of the sequence the target is dim and is barely distinguishable from the background. There is considerable overlap between the target and background histograms throughout, as is typical for sequences acquired under practical field conditions. Accurate histogram estimation is critical in such cases, since the accumulation of small errors can corrupt the target model and ultimately cause the track filter to lock onto background structure and fail. Improved histogram estimation was achieved by modeling the temporal evolution of the reference histogram in an adaptive Kalman filtering (AKF) framework in [30]. In [2], [40], the AKF measurement noise variance was estimated from the first frame and was assumed stationary, while the process noise variance was estimated online using covariance matching [41]. A robust Kalman filter was developed for template-based appearance learning in [25], where the process noise was assumed known and covariance matching was used to estimate the innovations variance.

#### D. Original Contributions

In this paper, we present a new histogram-based appearance learning algorithm where intensity histograms for both the target and background are updated in each frame by a bank of

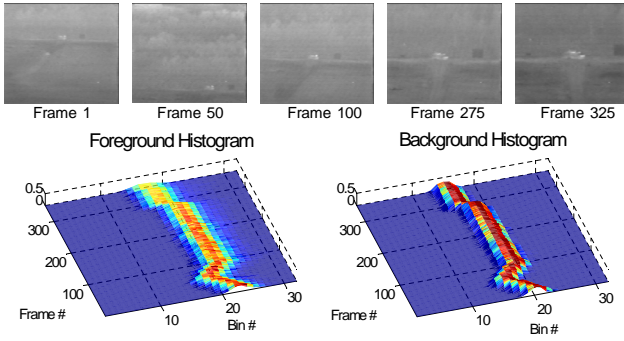


Fig. 2. Nonstationary evolution of target foreground/background in AMCOM sequence rng17\_01. Five raw IR frames are shown above time plots of the target histogram (left) and the local background histogram (right).

AKF's. For the first time in an appearance learning application, the unknown process and measurement noise variances are estimated simultaneously using the recently developed auto-covariance least-squares (ALS) method [42], [43]. In order to provide robustness, to accommodate strong ego-motion, and to provide flexibility in dealing with dynamic target size estimation, we adopt a particle filter-based tracker where the state vector gives the target position and magnification and the likelihood function depends on the adaptive appearance model. The proposed algorithm is able to estimate the target position in challenging IR imagery with an average error of less than 1.2 and 2 pixels respectively against the AMCOM and SENSIAC datasets, achieving sub-pixel accuracy in many cases. Estimation of the target magnification, which is normally under-treated in infrared tracking, is achieved with an average error of two to four pixels for both the AMCOM and SENSIAC sequences. We believe these results are among the best reported against the AMCOM sequences and among the best and earliest reported against the SENSIAC data.

The main contributions of this paper include application of the ALS covariance estimation method in visual target tracking, adaptation of the ALS method to block stationary system dynamics, development of a robust appearance learning algorithm based on a quad of dual foreground-background histograms, and integration of these techniques to achieve near sub-pixel tracking accuracy against the AMCOM and SENSIAC sequences. The new appearance learning and tracking techniques introduced here are distinct from those given in [2], [23], [30] in the use of a particle filter as opposed to the mean-shift algorithm and from those in [2], [25], [40] in the use of histogram-based appearance learning. The experiments in Section III demonstrate that the new ALS-based histogram learning outperforms traditional histogram similarity (HS)-based update methods [32], [33] and the previous AFK-based method in [30] where the covariance matching (COV) technique was used to estimate unknown noise parameters.

## II. HISTOGRAM-BASED APPEARANCE LEARNING

Let  $\mathbf{y}_k$  be a sequence of video frames acquired from a passive imaging sensor at discrete time instants  $k \in \mathbb{N}$ . For simplicity, we assume that there is a single object of interest, which could be, e.g., a target or a patch of background. Let

$\mathbf{g}_k = \{g_k^b\}_{b=1, \dots, N_b}$  be the observed normalized histogram of the object computed from the frame  $\mathbf{y}_k$ , where  $\sum_{b=1}^{N_b} g_k^b = 1$  and the histogram is discretized to  $N_b$  bins. Similarly, let  $\mathbf{f}_k = \{f_k^b\}_{b=1, \dots, N_b}$  be the reference histogram, which provides an idealized model of the object appearance at time  $k$ .

The objective of histogram learning is to estimate the present appearance model  $\mathbf{f}_k$  by incorporating the current observation  $\mathbf{g}_k$  into the previous appearance model  $\mathbf{f}_{k-1}$ . This is typically formulated as a time-varying linear filter

$$\mathbf{f}_k = \boldsymbol{\xi}_k \cdot \mathbf{g}_k + (1 - \boldsymbol{\xi}_k) \cdot \mathbf{f}_{k-1}, \quad (1)$$

where  $\mathbf{1}$  is a vector with all entries equal to one and “ $\cdot$ ” represents the Hadamard (or Schur) product. The vector  $\boldsymbol{\xi}_k = \{\xi_k^b\}_{b=1, \dots, N_b}$  controls the balance between the previous reference model  $\mathbf{f}_{k-1}$  and the new observation  $\mathbf{g}_k$ , where  $0 \leq \xi_k^b \leq 1$  is the time dependent filter coefficient for the  $b$ th histogram bin. Accurate tuning of  $\boldsymbol{\xi}_k$  is the key to effective appearance learning.

In this section we discuss three different learning techniques that share the form (1) and differ only in how  $\boldsymbol{\xi}_k$  is computed. The first is the traditional histogram similarity based method where all bins are updated with the same coefficient ( $\xi_k^b = \xi_k; b = 1, \dots, N_b$ ). We shall refer to this method as HS. After briefly reviewing the basic Kalman filter, we turn our attention to two AKF methods that use different approaches for estimating the process and measurement noise variances. The first, which we will call AKF<sub>cov</sub>, uses covariance matching where the same coefficient is applied to all bins. The second, which we refer to as AKF<sub>als</sub>, uses the recent ALS technique [42], [43] and maintains a separate coefficient  $\xi_k^b$  for each histogram bin.

### A. Histogram Similarity Method (HS)

In the widely used HS method, the coefficient vector  $\boldsymbol{\xi}_k$  in (1) is updated based on histogram similarity [33], [44]. All  $N_b$  entries of  $\boldsymbol{\xi}_k$  share a common value given by the metric

$$\xi_k = 1 - h(\mathbf{f}_{k-1}, \mathbf{g}_k), \quad (2)$$

where  $h$  is a normalized histogram similarity measure such as the Bhattacharyya coefficient [17]. In practice, however, we find that the histogram intersection defined by [44]

$$h(\mathbf{f}_{k-1}, \mathbf{g}_k) = \sum_{i=1}^{N_b} \min(f_{k-1}^i, g_k^i) \quad (3)$$

is more useful for quantifying histogram similarity in IR imagery. With (3), if the observed and reference histograms are nearly identical then  $h(\mathbf{f}_{k-1}, \mathbf{g}_k) \approx 1$  and  $\xi_k$  is small, implying that very little information from the observation will be incorporated into the learning process at time step  $k$ . Alternatively, if the two histograms are almost mutually exclusive then  $h(\mathbf{f}_{k-1}, \mathbf{g}_k) \approx 0$  and  $\xi_k \approx 1$ , implying that the new reference histogram will be heavily dependent on the observation and will largely discard the historical information contained in  $\mathbf{f}_{k-1}$ . Thus, the observation is weighted strongly when there is a sudden change in the object appearance. Note that the similarity metric (2), (3) depends on all  $N_b$  histogram bins and is scalar-valued, implying that a common

coefficient  $\xi_k$  is applied to all bins in the HS method. As with many dynamic appearance learning strategies, the HS method can potentially over adapt in the presence of strong measurement noise and/or rapidly evolving target signatures, causing track loss due to the target appearance model becoming corrupted with background information. Explicit outlier rejection algorithms were implemented in [30], [40] to mitigate this problem.

### B. Kalman Filtering

To reformulate appearance learning as a Kalman filtering problem, we model corresponding bins  $f_k^b$  and  $g_k^b$  from the reference and observed histograms in state space according to

$$f_{k+1}^b = f_k^b + w_k^b, \quad (4)$$

$$g_k^b = f_k^b + v_k^b, \quad (5)$$

where  $w_k^b$  and  $v_k^b$  are mutually uncorrelated process and measurement noises, both assumed zero-mean, white, and Gaussian with variances  $\sigma_{wb}^2(k)$  and  $\sigma_{vb}^2(k)$  that are time-varying in general. The Kalman filter state prediction and update equations for the system are given by

$$\text{State prediction: } \hat{f}_{k|k-1}^b = \hat{f}_{k-1}^b \quad (6)$$

$$\text{Covariance prediction: } p_{k|k-1}^b = p_{k-1}^b + \sigma_{wb}^2(k) \quad (7)$$

$$\text{Kalman gain: } K_k^b = \frac{p_{k|k-1}^b}{p_{k|k-1}^b + \sigma_{vb}^2(k)} \quad (8)$$

$$\text{Innovation: } r_k^b = g_k^b - \hat{f}_{k|k-1}^b \quad (9)$$

$$\begin{aligned} \text{State update: } \hat{f}_k^b &= \hat{f}_{k|k-1}^b + K_k^b r_k^b \\ &= K_k^b g_k^b + (1 - K_k^b) \hat{f}_{k-1}^b \end{aligned} \quad (10)$$

$$\text{Covariance update: } p_k^b = (1 - K_k^b) p_{k|k-1}^b. \quad (11)$$

There is a direct correspondence between (1) and (10), where the Kalman gain  $K_k^b$  in (10) may be associated with the coefficient  $\xi_k^b$  in (1); hence, with the Kalman filtering formulation we obtain  $\xi_k^b \equiv K_k^b$ .

The Kalman filter balances the relative contributions to appearance learning from the reference and observed data based on the estimated variances  $\sigma_{wb}^2(k)$  and  $\sigma_{vb}^2(k)$ . When  $\sigma_{wb}^2(k) \gg \sigma_{vb}^2(k)$ , for example, we have  $K_k^b \approx 1$  implying that the observation will be weighted much more heavily than the historical reference data. Under the linearity and Gaussianity assumptions applied here, the state estimates (6) and (10) are optimal in the minimum mean squared error sense.

However, computing the Kalman gains (8) requires knowledge of  $\sigma_{wb}^2(k)$  and  $\sigma_{vb}^2(k)$ , both of which are usually unknown in practice. This leads to the adaptive Kalman filter (AKF), which seeks to estimate the unknown noise variances on the fly. A brief overview of AKF methods was given in [41] and more recent surveys appear in [45], [46]. In [41], these techniques were broadly divided into four categories: Bayesian, maximum likelihood (ML), correlation, and covariance matching methods. The Bayesian method requires the evaluation of several difficult integrals and the ML method relies on equations that involve partial derivatives thereby making them both computationally expensive. The correlation

and covariance matching methods relate certain properties of the filter residues with the unknown noise processes using linear equations, which allows for easy representation and computations using simple matrix operations. For these reasons, in the following we focus on two different AKF-based appearance learning algorithms that rely on the covariance matching and correlation approaches.

### C. AKF: Covariance Matching (AKF<sub>cov</sub>)

Covariance matching techniques [41], [47] are based on the relationship that exists between the process and measurement noise variances and the autocorrelation of the innovations process (9). Since the innovations are observable, their autocorrelation can be estimated by an empirical sample variance under suitable ergodicity assumptions. Thus, if one of the two variances  $\sigma_{wb}^2(k)$  and  $\sigma_{vb}^2(k)$  is known, then the other one can be estimated by matching the empirically calculated innovations autocorrelation to its theoretical value. Here, we adopt the specific technique used in [2], [30], [40] where  $\sigma_{vb}^2(k)$  is known and  $\sigma_{wb}^2(k)$  is obtained by covariance matching.

It follows easily from (4)-(11) that the autocorrelation of the innovations process is given by [48, Section V.B]

$$E[r_k^b r_j^b] = [p_{k-1}^b + \sigma_{vb}^2(k) + \sigma_{wb}^2(k-1)] \delta(k-j), \quad (12)$$

where  $\delta(\cdot)$  is the Kronecker delta. With  $\sigma_{vb}^2(k)$  known and  $p_{k-1}^b$  given by (11), an obvious empirical approach for solving  $\sigma_{wb}^2(k-1)$  from (12) is to approximate  $E[(r_k^b)^2]$  by computing the sample variance of (9) over the last  $L_{\text{cov}}$  frames  $\mathbf{y}_{k-L_{\text{cov}}+1}, \dots, \mathbf{y}_k$ . However, because the process noise could be time varying in general, there is a delicate tradeoff between choosing  $L_{\text{cov}}$  large enough to obtain statistically significant estimates while simultaneously choosing  $L_{\text{cov}}$  small enough to track nonstationary changes in  $\sigma_{wb}^2(k)$ .

In appearance learning for visual target tracking, this problem has been addressed previously by assuming identical statistics across variables in order to increase the sample size to larger than  $L_{\text{cov}}$  while still sampling from only the  $L_{\text{cov}}$  most recent frames. In [30], it is assumed that  $\sigma_{vb}^2(k)$  is independent of both  $b$  and  $k$  and that  $\sigma_{wb}^2(k)$  is independent of  $b$ , so that all  $N_b$  bins of the histogram in each frame share identical noise statistics. The innovations sample variance may then be computed across bins as well as over time. The same assumptions on  $\sigma_{vb}^2(k)$  are made for the template-based appearance model of [40], where  $b$  indexes pixels in the template rather than bins in the histogram. By assuming a common value  $\sigma_{wb}^2(k)$  for all template pixels in the current frame, the innovations sample variance can be averaged across both pixels and time. A similar strategy was employed in [2] with the principal difference that  $\sigma_{vb}^2(k)$  was assumed time varying and estimated by an auxiliary algorithm independent of the covariance matching. Similar covariance matching was used to estimate the scale matrix in [25].

To formulate this class of covariance matching algorithms in our present setup, we assume that  $\sigma_{vb}^2(k)$  is independent of both  $k$  and  $b$  and that  $\sigma_{wb}^2(k)$  is independent of  $b$  (as in [30], [40]). Let  $\mathcal{B}$  be the set of nonzero histogram bins and estimate

$E[(r_k^b)^2]$  with the sample variance

$$\hat{C}_r(k) = \frac{1}{|\mathcal{B}|L_{\text{cov}}} \sum_{i=0}^{L_{\text{cov}}-1} \sum_{b \in \mathcal{B}} (r_{k-i}^b)^2. \quad (13)$$

Under these assumptions  $p_{k-1}^b$  is independent of  $b$ . Thus, we arbitrarily choose  $p_{k-1}^1$  and use (13) in (12) to obtain the approximate solution

$$\sigma_{wb}^2(k-1) \approx \hat{C}_r(k) - \sigma_{vb}^2(k) - p_{k-1}^1. \quad (14)$$

As in [30], [40], the initialization at  $k = 1$  is given by

$$\sigma_{vb}^2(k) = \frac{1}{2} \hat{C}_r(1) \quad \forall b, k; \quad p_0^b = \frac{1}{2} \hat{C}_r(1) \quad \forall b, \quad (15)$$

which implies  $\sigma_{wb}^2(0) = 0$ . We refer to this algorithm as  $\text{AKF}_{\text{cov}}$  and use it in the following as a baseline for comparison with the  $\text{AKF}_{\text{als}}$  technique given in the next section.

#### D. AKF: Autocovariance-Based Least Squares ( $\text{AKF}_{\text{als}}$ )

The ideal expression (12) for the innovations autocorrelation holds when there are no modeling errors and the filter gains  $K_k^b$  in (8) are optimal. However, if the process and measurement noise variances are unknown then the gains will be suboptimal and the innovations process will generally exhibit a nontrivial correlation structure. The main idea of autocovariance based methods is to exploit any observed nonzero correlations at lags other than zero to obtain solutions for the unknown noise variances and/or the optimal gains. Pioneering work in this area was given by Mehra in [41], [49] where the residual autocorrelation was used for adaptive Kalman filtering. Mehra's method involves a three-step iterative process where a Lyapunov-type equation must be solved at every time step. Under the assumption that the process and measurement noises are wide sense stationary (WSS), Carew and Bélanger [50] developed an improved algorithm that estimates the optimal Kalman gains directly using one matrix inversion and several matrix multiplications, eliminating the need to estimate the process and measurement noise variances explicitly and avoiding the requirement to iteratively solve the Lyapunov equation associated with Mehra's method. Neethling and Young [51] introduced a related weighted least squares technique that improves the statistical efficiency of the methods in [41], [49], [50] and incorporates a side constraint to guarantee positive semi-definite (PSD) estimates for the unknown noise variances.

Recently, Odelson, *et al.*, developed a new Autocovariance Least Squares (ALS) method capable of providing PSD estimates for both the process and measurement noise variances simultaneously [42], [43]. In addition, the ALS variance estimates are more stable than those delivered by Mehra's method and converge asymptotically to the optimal values with increasing sample size. However, the proof of convergence given in [42], [52] depends explicitly on assumptions that the system is time invariant and that the process and measurement noises are WSS (extension to a time varying system with WSS noises was given in [53]). The ALS algorithm in [42] is primarily meant for identifying the system noise properties in an offline learning process under WSS assumptions. First, the

filter innovations are obtained from the observations using a suboptimal Kalman gain over an extended period of time. Then the autocovariance structure of these innovations is used to reliably estimate the noise variances. Once the noise variances are known, the optimal Kalman gain can be determined and applied for filtering during run time using the standard Kalman filtering equations (6)-(11).

For appearance learning, our interest in this paper is primarily in real-time, online scenarios where, for the first time, we consider application of the ALS method under the much weaker assumption that the noise variances  $\sigma_{wb}^2(k)$  and  $\sigma_{vb}^2(k)$  are only block stationary. In order to extend the ALS method to this case, we consider the evolution of the target appearance to be a piecewise stationary process with non-stationary transitions. The piecewise stationarity assumption can be justified by the high frame rate of the imaging sensor compared to the rate at which the target appearance changes. Such assumptions are common, *e.g.*, in the context of audio and video compression [54]–[60]. The nonstationary characteristics of  $\sigma_{wb}^2(k)$  and  $\sigma_{vb}^2(k)$  directly correlate with the rate at which the target appearance and sensor noise are changing. The piecewise stationary formulation allows us to apply the ALS algorithm to each stationary block individually and thereby allows us to adapt to the varying nature of the target appearance histogram over time. In effect, we adapt the filter gain  $\xi_k^b$  at the end of each stationary block depending on the observed variation trend in that block. This raises the issue of determining the block boundaries. Most existing methods that determine the block intervals require *a priori* knowledge of the observations; since this is not the case in our real-time application, we consider equal length blocks. We study the effect of block size by performing experiments using the ALS method on a simulated nonstationary system in Section II-E.

In this section, we extend the ALS method for application to a piecewise stationary process in the context of histogram-based appearance learning, which we refer to as  $\text{AKF}_{\text{als}}$  in this paper. As before, the state model is given by (4) and (5). We assume that  $w_k^b$  and  $v_k^b$  are mutually uncorrelated and that  $\sigma_{wb}^2(k)$  and  $\sigma_{vb}^2(k)$  depend on  $b$  and  $k$  and are piecewise constant. With this setup, the noise statistics are generally different for each bin of the histogram and there is a separate coefficient  $\xi_k^b$  for each  $b \in [1, N_b]$ . The size of each piecewise stationary block is assumed to be  $N_d$  frames. We also define a block index  $p$ , where the  $p$ th block contains frames  $\mathbf{Y}(p) = \{\mathbf{y}_k | k \in \mathbb{K}(p)\}$  with

$$\mathbb{K}(p) = \{k | (p-1)N_d + 1 \leq k \leq pN_d\}. \quad (16)$$

Using this framework, we update the estimated noise variances of the appearance histogram corresponding to each bin at the end of every block. In effect, we are adapting the filter gain (learning rate) for the current block based on the observed variations in the preceding block.

We now briefly present the least squares formulation to determine the system noise variances for a histogram bin. For the remainder of the section, we drop the bin index  $b$  for brevity. We assume that the asymptotic Kalman gain  $K_{p-1}$  estimated from the previous block is available. Given  $K_{p-1}$ , the state estimates in (10) for all frames in  $\mathbb{K}(p)$  are given by

$\hat{f}_k = \hat{f}_{k|k-1} + K_{p-1}r_k$ . The error in the predicted state (6) is defined by  $\varepsilon_k = f_k - \hat{f}_{k|k-1}$ . Then, for all frames  $k \in \mathbb{K}(p)$ , this prediction error along with the innovation (9) can be formulated together in a state model according to [42], [43]

$$\begin{aligned}\varepsilon_{k+1} &= \overbrace{(1 - K_{p-1})}^{\bar{a}_p} \varepsilon_k + \overbrace{\begin{bmatrix} 1 & -K_{p-1} \end{bmatrix}}^{\bar{\mathbf{g}}_p} \overbrace{\begin{bmatrix} w_k \\ v_k \end{bmatrix}}^{\bar{\mathbf{w}}_p}, \\ r_k &= \varepsilon_k + v_k.\end{aligned}\quad (17)$$

The ALS method aims to observe the filter innovations and exploit any observed nonzero correlations at different lags to obtain solutions for the unknown noise variances and/or the optimal gains. The autocorrelation of the innovations in the  $p$ th block at any lag  $j$  is given by

$$\mathcal{C}_j(p) = E[r_k r_{k+j}]; \quad 0 \leq j < L_{\text{als}}, \quad (19)$$

where  $k, k+j \in \mathbb{K}(p)$  and  $L_{\text{als}} < N_d$  is the order of the autocorrelation lags we consider in formulating the ALS problem. We assume  $E[\varepsilon_0] = 0$  and  $\text{cov}(\varepsilon_0) = \pi_0$  and define

$$\bar{\mathbf{Q}}_p = E[\bar{\mathbf{w}}_p \bar{\mathbf{w}}_p^T] = \begin{bmatrix} \sigma_w^2(pN_d) & 0 \\ 0 & \sigma_v^2(pN_d) \end{bmatrix}, \quad (20)$$

$$\bar{\mathbf{x}}_p = E[\bar{\mathbf{w}}_p v_k] = \begin{bmatrix} 0 \\ \sigma_v^2(p) \end{bmatrix} \quad (21)$$

for  $k \in \mathbb{K}(p)$ . Note in (21) that although  $E[\bar{\mathbf{w}}_p v_k]$  contains the time index  $k$ , this expectation is constant over  $\mathbb{K}(p)$  due to the piecewise stationarity assumption.

In the interest of clarity and to illustrate the form of the relevant relations, we assume  $L_{\text{als}} = 3$  in the following; generalization to other  $L_{\text{als}}$  is straightforward. The least squares estimation problem is formulated in terms of an autocovariance matrix  $\mathbf{R}_p(L_{\text{als}})$  that, for  $L_{\text{als}} = 3$  and  $k, k+1, k+2 \in \mathbb{K}(p)$ , is given by

$$\mathbf{R}_p(3) = E \begin{bmatrix} (r_k)^2 & r_k r_{k+1} & r_k r_{k+2} \\ r_k r_{k+1} & (r_{k+1})^2 & r_{k+1} r_{k+2} \\ r_k r_{k+2} & r_{k+1} r_{k+2} & (r_{k+2})^2 \end{bmatrix}. \quad (22)$$

The individual elements of  $\mathbf{R}_p(3)$  are functions of  $\pi_0, \bar{a}_p, \bar{\mathbf{g}}_p, \bar{\mathbf{Q}}_p$  and  $\bar{\mathbf{x}}_p$ . Let “vec” be the vectorization operator that transforms a matrix into a vector by stacking the columns upon one another. The vectorization of  $\mathbf{R}_p(3)$  is given by

$$\begin{aligned}\text{vec}[\mathbf{R}_p(3)] &= (\boldsymbol{\theta}_p \otimes \boldsymbol{\theta}_p) \pi_0 \\ &\quad + \boldsymbol{\Gamma}_p \otimes \boldsymbol{\Gamma}_p \text{vec}[\mathbf{I}_3] \text{vec}[\bar{\mathbf{g}}_p \bar{\mathbf{Q}}_p \bar{\mathbf{g}}_p^T] \\ &\quad + (\boldsymbol{\Psi}_p \oplus \boldsymbol{\Psi}_p + \mathbf{I}_3) \text{vec}[\mathbf{I}_3] \sigma_v^2(p),\end{aligned}\quad (23)$$

where  $\mathbf{I}_n$  denotes the  $n \times n$  identity matrix,  $\otimes$  denotes the Kronecker product,  $\oplus$  denotes Kronecker sum, and the matrices  $\boldsymbol{\theta}_p, \boldsymbol{\Gamma}_p$  and  $\boldsymbol{\Psi}_p$  are given by

$$\boldsymbol{\theta}_p = \begin{bmatrix} 1 \\ \bar{a}_p \\ \bar{a}_p^2 \end{bmatrix}; \boldsymbol{\Gamma}_p = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ \bar{a}_p & 1 & 0 \end{bmatrix}; \boldsymbol{\Psi}_p = -K_{p-1} \boldsymbol{\Gamma}_p \mathbf{I}_3. \quad (24)$$

Using the Lyapunov equation to eliminate the  $\pi_0$  term in (23), one obtains

$$\overbrace{\text{vec}[\mathbf{R}_p(3)]}^{\mathcal{R}_p(3)} = \overbrace{\begin{bmatrix} \mathbf{d}_p & \mathbf{d}_p K_{p-1}^2 + (\boldsymbol{\Psi}_p \oplus \boldsymbol{\Psi}_p + \mathbf{I}_9) \text{vec}[\mathbf{I}_3] \end{bmatrix}}^{\mathcal{A}_p} \overbrace{\begin{bmatrix} \sigma_w^2(p) \\ \sigma_v^2(p) \end{bmatrix}}^{\mathbf{x}_p}, \quad (25)$$

where  $K_{p-1}$  is scalar,  $\mathcal{R}_p(3)$  is  $9 \times 1$ ,  $\mathcal{A}_p$  is  $9 \times 2$ ,  $\mathbf{x}_p$  is  $2 \times 1$ , and  $\mathbf{d}_p$  is a  $9 \times 1$  vector defined by

$$\mathbf{d}_p = (\boldsymbol{\theta}_p \otimes \boldsymbol{\theta}_p)(1 - \bar{a}_p^2)^{-1} + (\boldsymbol{\Gamma}_p \otimes \boldsymbol{\Gamma}_p) \text{vec}[\mathbf{I}_3]. \quad (26)$$

$\mathcal{R}_p(3)$  may also be represented in terms of the autocorrelation terms defined in (19) according to

$$\mathcal{R}_p(3) = \text{vec} \begin{bmatrix} \mathcal{C}_0(p) & \mathcal{C}_1(p) & \mathcal{C}_2(p) \\ \mathcal{C}_1(p) & \mathcal{C}_0(p) & \mathcal{C}_1(p) \\ \mathcal{C}_2(p) & \mathcal{C}_1(p) & \mathcal{C}_0(p) \end{bmatrix}. \quad (27)$$

Provided that the innovations process is reasonably locally ergodic, the quantities  $\mathcal{C}_j(p)$  in (27) may be estimated by

$$\hat{\mathcal{C}}_j(p) = \frac{1}{N_d - j} \sum_{i=(p-1)N_d+1}^{pN_d-j} r_i r_{i+j}. \quad (28)$$

We define an estimated vectorized correlation matrix  $\hat{\mathcal{R}}_p(3)$  by replacing the theoretical correlations  $\mathcal{C}_j(p)$  in (27) with the empirical estimates  $\hat{\mathcal{C}}_j(p)$  given by (28). From this definition and (25), we write

$$\mathcal{A}_p \mathbf{x}_p = \hat{\mathcal{R}}_p(3). \quad (29)$$

The expression (29) forms the core of the ALS method: it relates the observed correlations contained in  $\hat{\mathcal{R}}_p(3)$  and defined in (28) to the desired variances  $\sigma_w^2(p)$  and  $\sigma_v^2(p)$  contained in  $\mathbf{x}_p$ . Also note that  $\mathcal{A}_p$  is dependent only on the asymptotic Kalman gain  $K_{p-1}$  from the previous block. Thus, the least squares problem for the unknown noise variances  $\sigma_w^2(p)$  and  $\sigma_v^2(p)$  can be expressed as

$$\Phi_p = \min_{\sigma_w^2(p), \sigma_v^2(p)} \left\| \mathcal{A}_p \begin{bmatrix} \sigma_w^2(p) \\ \sigma_v^2(p) \end{bmatrix} - \hat{\mathcal{R}}_p(3) \right\|^2 \quad (30)$$

subject to  $\sigma_w^2(p), \sigma_v^2(p) \geq 0$ . The positive semi-definite requirements on  $\sigma_w^2(p)$  and  $\sigma_v^2(p)$  are enforced by appending a logarithmic barrier function to (30), resulting in

$$\begin{aligned}\Phi_p &= \min_{\sigma_w^2(p), \sigma_v^2(p)} \left\| \mathcal{A}_p \begin{bmatrix} \sigma_w^2(p) \\ \sigma_v^2(p) \end{bmatrix} - \hat{\mathcal{R}}_p(3) \right\|^2 \\ &\quad - \mu \log[\sigma_w^2(p) \sigma_v^2(p)],\end{aligned}\quad (31)$$

where  $\mu$  is the barrier parameter. The least squares problem (31) has been shown to be convex and can be solved using a Newton recursion [42]. Pseudo-code to implement this AKF<sub>als</sub> algorithm for a single bin of the histogram is given in Table I.



TABLE I  
PSEUDO-CODE TO IMPLEMENT  $\text{AKF}_{\text{als}}$  FOR A SINGLE BIN OF THE  
APPEARANCE HISTOGRAM DURING THE  $p$ TH TIME BLOCK.

---

For  $k = (p-1)N_d + 1$  to  $pN_d$

1. Predict bin value  $\hat{f}_{k|k-1} = \hat{f}_{k-1}$ .
2. Acquire observation  $g_k$  based on tracker output.
3. Compute innovation  $r_k = g_k - \hat{f}_{k|k-1}$ .
4. Update bin value  $\hat{f}_k = \hat{f}_{k|k-1} + K_{p-1}r_k$ .

End

5. Find  $\hat{\mathcal{R}}_p(L_{\text{als}})$  from  $\hat{\mathcal{C}}_j(p)$  for  $0 \leq j \leq L_{\text{als}} - 1$  using (28)
6. Determine  $\mathcal{A}_p$  using (25) to set up the ALS problem (29).
7. Perform the optimization in (31) to obtain  $\sigma_v^2(p)$  and  $\sigma_w^2(p)$ .
8. Compute asymptotic Kalman gain  $K_p$  from the estimated noise variances for use in the next block.

---

### E. Numerical simulations

Having extended the ALS method to the piecewise stationary case, we perform two numerical experiments on simulated data. The first one compares the noise variance estimation capability of  $\text{AKF}_{\text{cov}}$  and  $\text{AKF}_{\text{als}}$  on a system with WSS noise characteristics. The second examines the performance of the proposed piecewise stationary ALS method against piecewise stationary and more general nonstationary system dynamics.

1) *Comparison between  $\text{AKF}_{\text{als}}$  and  $\text{AKF}_{\text{cov}}$* : The objective of this experiment is to estimate the unknown noise covariance matrices from simulated data using  $\text{AKF}_{\text{cov}}$  and  $\text{AKF}_{\text{als}}$ . Consider a system of the form

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \quad (32)$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k, \quad (33)$$

where  $\mathbf{w}_k$  and  $\mathbf{v}_k$  are zero mean, iid Gaussian noise processes with fixed covariances  $\mathbf{Q}$  and  $\mathbf{R}$ , respectively. Let

$$\mathbf{A} = \begin{bmatrix} 0.1 & 0 & 0.1 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.3 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & -0.1 & 0.2 \\ -0.2 & 1 & 0 \\ 0 & -0.4 & 1 \end{bmatrix},$$

$$\mathbf{Q} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.75 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 0.75 \end{bmatrix}.$$

During the estimation process, the diagonal elements of the estimates of  $\mathbf{Q}$  and  $\mathbf{R}$  were initialized with random values uniformly distributed between zero and one. The asymptotic filter gain for the initialized noise covariances was then computed. This gain was used for filtering against 5000 data points to obtain innovations that were used, along with the initial estimates of  $\mathbf{Q}$  and  $\mathbf{R}$ , by the  $\text{AKF}_{\text{cov}}$  and  $\text{AKF}_{\text{als}}$  methods to estimate the unknown noise covariances. Results from repeating the simulation 200 times are shown in Fig. 3, where each point corresponds to the estimate from a single trial. It is observed that  $\text{AKF}_{\text{als}}$  produces estimates that are PSD and more precise than those delivered by  $\text{AKF}_{\text{cov}}$ . The estimates produced by  $\text{AKF}_{\text{cov}}$  seem to depend on the initial values of the unknowns. Since  $\text{AKF}_{\text{cov}}$  assumes that at least one of the noise covariances is known *a priori*, an erroneous initial value can greatly distort the estimation. Further, there is no guarantee that the estimates (14) are PSD, as seen by the occasional negative estimates of the  $\text{AKF}_{\text{cov}}$  method in Fig. 3. Unlike

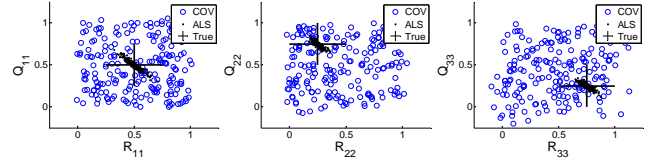


Fig. 3. Diagonal elements of the noise covariance matrices  $\mathbf{Q}$  and  $\mathbf{R}$  as estimated by  $\text{AKF}_{\text{cov}}$  and  $\text{AKF}_{\text{als}}$  for WSS system dynamics.

$\text{AKF}_{\text{cov}}$ ,  $\text{AKF}_{\text{als}}$  (1) estimates both process and observation noise parameters simultaneously, (2) formulates a least squares problem based on multiple constraints obtained by considering the autocorrelation of the innovations at different lags, and (3) enforces PSD constraints on the estimates. The use of multiple constraints in the least squares solution greatly diminishes the effect of erroneous initial values.

2) *Piecewise treatment of non-stationary systems by  $\text{AKF}_{\text{als}}$* : Here, we examine the performance of  $\text{AKF}_{\text{als}}$  and its inherent block stationarity assumptions against the linear state model (32), (33) for the case of nonstationary noise processes  $\mathbf{w}_{k-1}$  and  $\mathbf{v}_k$  with diagonal covariance matrix entries that exhibit jump transitions and linear ramps. Let

$$\mathbf{A} = \begin{bmatrix} 0.9 & 0 & 0.7 \\ 0 & 0.95 & 0 \\ 0 & 0 & 0.7 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & -0.1 & 0.2 \\ -0.2 & 1 & 0 \\ 0 & -0.4 & 1 \end{bmatrix},$$

and let  $\mathbf{w}_k$  and  $\mathbf{v}_k$  be zero mean, iid Gaussian noise processes with time varying diagonal covariance matrices  $\mathbf{Q}$  and  $\mathbf{R}$  having main diagonal entries given by the dotted (blue) lines in Fig. 4. As indicated in the figure, the noise covariances are block stationary during the first portion of each simulation and increase or decrease linearly with small-scale additive noise during the second portion. The transition times between these characteristics for all six diagonal covariance matrix entries are mutually independent.

The objective is to estimate the six unknown covariances using the  $\text{AKF}_{\text{als}}$  algorithm developed in Section II-D. In the absence of any *a priori* knowledge about the transition times between piecewise stationary and linear characteristics in the noise variances, we set the block length  $N_d$  in the  $\text{AKF}_{\text{als}}$  algorithm to a constant. Choosing  $N_d$  small results in a paucity of data points being available to perform statistically significant least squares estimation, whereas choosing  $N_d$  large limits the ability of the algorithm to adapt to the nonstationary changes. The experiment is designed to study the performance of  $\text{AKF}_{\text{als}}$  as a function of the chosen block size  $N_d$ .

The estimates of the diagonal elements of  $\mathbf{Q}$  and  $\mathbf{R}$  are initialized with random values distributed uniformly between zero and one. The asymptotic Kalman gain corresponding to this initialization is used for filtering over the first block of length  $N_d$  to obtain innovations. These innovations are then used to formulate the least squares problem (31), the solution of which yields estimates for the six unknown noise covariances and an asymptotic Kalman gain  $K_1$ . In an offline application, this Kalman gain could be used to re-process the first block. For a real-time implementation, however, we instead use the asymptotic gain  $K_1$  obtained from the first

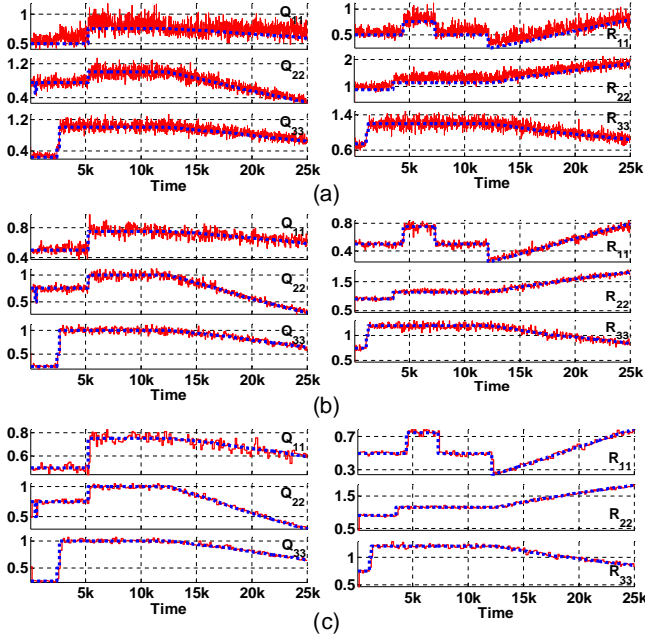


Fig. 4. Simulation of  $\text{AKF}_{\text{als}}$  against nonstationary noise statistics for three different block sizes. The dotted (blue) lines give the true values of the main diagonal entries of the process noise covariance matrix  $\mathbf{Q}$  (left column) and measurement noise covariance matrix  $\mathbf{R}$  (right column). The  $\text{AKF}_{\text{als}}$  covariance estimates are shown as solid (red) lines for (a)  $N_d = 15$ , (b)  $N_d = 45$ , and (c)  $N_d = 135$ .

block to process the data in the second block. This approach is effective for achieving real-time performance provided that the jump transitions are not too large and the ramp characteristics are not too steep. The procedure is repeated recursively with the gain  $K_{p-1}$  being used to process the data in block  $\mathbf{Y}(p)$  and generate innovations.

Since the number of constraints in the least squares problem should be larger than the number of unknowns (six in this case), we set the number of autocorrelation lags considered by the  $\text{AKF}_{\text{als}}$  algorithm to  $L_{\text{als}} = 10$ . We performed simulations against the covariances shown in Fig. 4 with block sizes  $N_d = 15, 45$ , and  $135$ , where 100 trials with different random initializations were run for each block size. The average estimated covariance values for the three different block sizes are shown as solid (red) lines in Fig. 4(a), (b), and (c). It is shown that a small block size ( $N_d = 15$ ) affords the opportunity to adapt quickly to abrupt nonstationary changes in the dynamics, but the estimation errors are generally large due to limited observations in each block. With the largest block size ( $N_d = 135$ ), the algorithm is slower in adapting to nonstationary changes, especially those that occur in the middle of a block, but the estimation errors are generally much smaller than with the small block size. Additionally, as the block size increases, the median error decreases and the probability of a large estimation error diminishes. Overall, we find that  $\text{AKF}_{\text{als}}$  can cope reasonably well with both the jump and ramp nonstationarities depending on the block size. These results show that the block-based ALS method that makes a piecewise stationary assumption can estimate the system noise covariances without manual filter tuning.

## F. Dual foreground-background appearance model

We present a target model that involves the local statistics of both the target and its surrounding background, as shown in Fig. 5. The use of background for tracking was discussed previously in [35]–[38]. In these methods, target tracking is performed on an intermediate classification image called a confidence map [35], a likelihood image [36], or a weighted image [37] where each pixel is assigned a probability of belonging to background or foreground. Here we have a different point of view using background for target modeling. Our target model is motivated by the “hit-and-miss” morphological transform that uses both foreground and background for object detection. In practice, the background information is found to be of great utility in localizing the target and determining its size. Specifically, the proposed target model involves four histograms to represent local statistics.

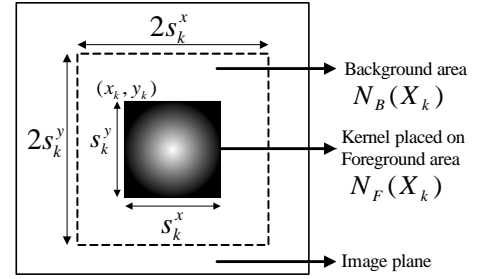


Fig. 5. Foreground region  $N_F(\mathbf{x}_k)$  with overlapped kernel and background area  $N_B(\mathbf{x}_k)$  defined based on  $\mathbf{x}_k = [x_k, y_k, s_k^x, s_k^y]$ .

Let  $\mathbf{x}_k = [x_k, y_k, s_k^x, s_k^y]$  be the state to be estimated during target tracking, where  $(x_k, y_k)$  and  $(s_k^x, s_k^y)$  are the position (top-left corner) and size of the target area in pixels. As shown in Fig. 5, the target appearance, denoted by  $\mathbf{G}(\mathbf{x}_k)$ , is composed of four histograms: the foreground/background intensity  $\mathbf{g}_A(\mathbf{x}_k)/\mathbf{g}_B(\mathbf{x}_k)$  and foreground/background local standard deviation (stdev)  $\mathbf{g}_C(\mathbf{x}_k)/\mathbf{g}_D(\mathbf{x}_k)$ , which are extracted from  $y_k$  by using the kernel-based method in [14], [15], [61]. Given  $\mathbf{x}_k$ , the candidate region is characterized by  $\mathbf{G}(\mathbf{x}_k)$  defined by

$$\mathbf{G}(\mathbf{x}_k) = \{\mathbf{g}_A(\mathbf{x}_k), \mathbf{g}_B(\mathbf{x}_k), \mathbf{g}_C(\mathbf{x}_k), \mathbf{g}_D(\mathbf{x}_k)\}. \quad (34)$$

A reference target model learned from previous frames is also available that is composed of four histograms, *i.e.*,  $\mathbf{F}_{k-1} = \{\mathbf{f}_{A,k-1}, \mathbf{f}_{B,k-1}, \mathbf{f}_{C,k-1}, \mathbf{f}_{D,k-1}\}$ . This reference model is updated online and used to evaluate any given candidate area in frame  $k$  represented by  $\mathbf{G}(\mathbf{x}_k)$  as

$$D(\mathbf{G}(\mathbf{x}_k), \mathbf{F}_{k-1}) = \sum_{z \in Z} v_z \cdot d(\mathbf{g}_z(\mathbf{x}_k), \mathbf{f}_{z,k-1}), \quad (35)$$

where  $Z = \{A, B, C, D\}$  and  $d$  is defined in (3);  $v_z$  is used to adjust the significance of the four histograms. Here, all four histograms are given equal importance. We develop a particle filter-based target tracking algorithm that uses this appearance model in conjunction with AKF-based appearance learning given in Table I as well as two dynamic models, one each for the position and size. The detailed tracking algorithm can be found in [15].



TABLE II  
LIST OF SEQUENCES USED IN EXPERIMENTS. TOP: AMCOM DATASET  
AND BOTTOM: SENSIAC DATASET

Sequences	Frame			Size	
	Starting frame	Ending frame	Length	Starting size	Ending size
LW-15-NS	21	270	250	5 x 8	16 x 16
LW-17-01	1	350	350	5 x 8	16 x 29
LW-21-15	236	635	400	3 x 4	10 x 10
LW-14-15	1	225	225	4 x 5	23 x 19
LW-22-08	51	300	250	5 x 8	17 x 24
LW-20-18	121	420	300	4 x 7	10 x 17
LW-18-17	1	190	190	5 x 9	11 x 25
LW-19-06	41	260	220	3 x 4	6 x 11
MW-14-10	1	450	450	6 x 11	12 x 28
LW-20-04	11	360	350	3 x 4	12 x 15
1925-0001	350	549	200	12 x 42	12 x 40
1925-0002	0	399	400	14 x 22	12 x 34
1925-0006	499	698	200	14 x 24	14 x 42
1925-0009	150	549	400	18 x 54	18 x 32
1925-0012	0	199	200	16 x 46	18 x 38
1927-0001	100	499	400	10 x 22	8 x 30
1927-0002	0	399	400	10 x 20	10 x 22
1927-0005	0	499	500	12 x 26	12 x 36
1927-0009	100	499	300	14 x 38	14 x 22
1927-0011	0	499	500	12 x 32	12 x 34

TABLE III  
DESCRIPTION AND VALUE OF THE EXPERIMENTAL PARAMETERS

Variables	Description	AMCOM	SENSIAC
$N_b^{(1)}$	bin number of the intensity histogram	32	32
$N_b^{(2)}$	bin number of the stdev histogram	16	16
$L_{cov}$	number of frames used for AKF <sub>cov</sub> in (13)	3	10
$N_d$	block size in frames	7	7
$L_{als}$	number of autocorrelation lags	5	2
$N_p$	number of particles used for tracking	200	100

### III. EXPERIMENTAL RESULTS

We tested the three histogram learning techniques along with the tracking algorithm presented in Section II against 10 sequences in each of the AMCOM and SENSIAC datasets. In order to represent the target appearance with a reasonable number of histogram bins, we performed contrast enhancement on the images from the SENSIAC dataset and down-sampled them to 8 bits. Further, the foreground intensity histogram for the SENSIAC dataset only includes pixels greater than 100 to maintain a good histogram structure. The metadata associated with both datasets provides ground truth for the target position, size and type, which are used to evaluate performance of the three appearance learning algorithms, HS, AKF<sub>cov</sub> and AKF<sub>als</sub>. The sequences selected for experiments from both datasets are enumerated in Table II. These sequences exemplify many of the important typical challenges of practical IR sequences, including poor target visibility, strong egomotion, small targets, significant pose variations and size variations, dust clouds, strong clutter and background noise, etc.

#### A. Experimental setup

Three appearance learning algorithms, namely HS [33], [44], AKF<sub>cov</sub> [30] and the proposed AKF<sub>als</sub>, are integrated with the same tracking algorithm for a fair comparison. All of them share the same linear histogram filtering form defined

in (1). HS determines  $\xi_k$  according to histogram similarity, while AKF<sub>cov</sub> and AKF<sub>als</sub> use the Kalman gain. Detailed parameterizations of the three algorithms are listed in Table III. In practice, AKF-based appearance learning algorithms were applied only to the two intensity histograms ( $f_A$  and  $f_B$ ). Because the dynamics of the stdev histograms do not have a well-defined structure, the stdev histograms ( $f_C$  and  $f_D$ ) in all cases were updated using the HS method. We also compare the performance of an alternative target representation using the covariance descriptor [62] that also supports online appearance updates. In addition to the tracking errors, we adopt an overlap metric proposed in [63] to quantify the degree of overlap between the track gate and the actual target area. Let  $A$  and  $B$  represent the track gate and the ground-truth bounding box respectively; then the overlap ratio  $\zeta$  is defined as

$$\zeta = \frac{\#(A \cap B) \times 2}{\#(A) + \#(B)}, \quad (36)$$

where  $\#$  is the number of pixels.

#### B. Experimental Analysis

The three algorithms (50 Monte Carlo runs each) were evaluated on 20 IR sequences from the AMCOM and SENSIAC datasets and compared numerically in terms of their appearance learning performance (Fig. 6), the overlap metric  $\zeta$  (Fig. 7) and the tracking error (Fig. 8 and Table V).

1) *Appearance learning*: Fig. 6 shows the histogram learning results for six AMCOM sequences, where it can be observed that the results of AKF<sub>als</sub> closely match the ground truth. Closer examination reveals that HS and AKF<sub>cov</sub> result in histograms that slowly deviate or “drift” from the ground truth. This is clearly evident in Fig. 6(c), where the intensity variations in the latter part of the sequence (around frame 300) are not captured by HS and AKF<sub>cov</sub>. Therefore, the tracker includes a large portion of the background in the track gate as seen in frames 320, 360 of the third sequence in Fig. 8 (a).

2) *Overlap metric*: Improvements in appearance learning against three AMCOM sequences are further demonstrated by the overlap metric in Fig. 7(a),(b), which compare  $\zeta_{als}$ ,  $\zeta_{cov}$  and  $\zeta_{HS}$  pairwise. For example, the improvement of AKF<sub>als</sub> over AKF<sub>cov</sub> or HS is demonstrated by observing that most data points are above the diagonal lines. Comparable results for AKF<sub>als</sub> and AKF<sub>cov</sub> against sequence LW-22-08 are also shown by the similar appearance learning performance in Fig. 6(e), where the histogram-based appearance model lacks strong modes and has widespread and small bin values. Average values of  $\zeta$  corresponding to the different algorithms against the two datasets are given in Table IV. AKF<sub>als</sub> has the largest values, indicating its superior target tracking performance compared to the other two algorithms.

3) *Tracking error*: Table V provides quantitative tracking performance results against the AMCOM and SENSIAC datasets. In most cases, AKF<sub>als</sub> achieves the smallest errors in terms of both position and size. The HS approach loses the target track in sequences LW-20-18 (6 runs) and LW-19-06 (2 runs), as indicated by the large errors. AKF<sub>cov</sub> also loses the target tracks in sequence LW-20-18 (1 run) due to the high

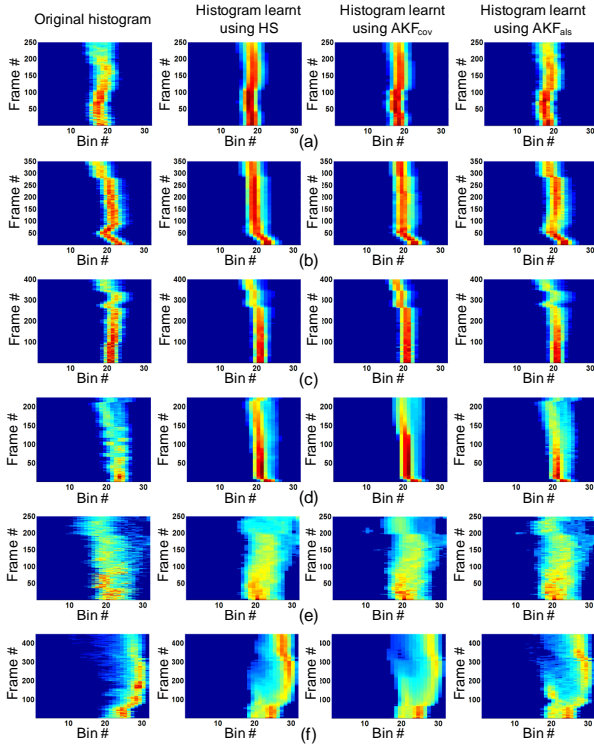


Fig. 6. Comparison of appearance learning for six AMCOM sequences using three methods: (a) LW-15-NS (b) LW-17-01 (c) LW-21-15 (d) LW-14-15 and (e) LW-22-08 and (f) MW-14-10.

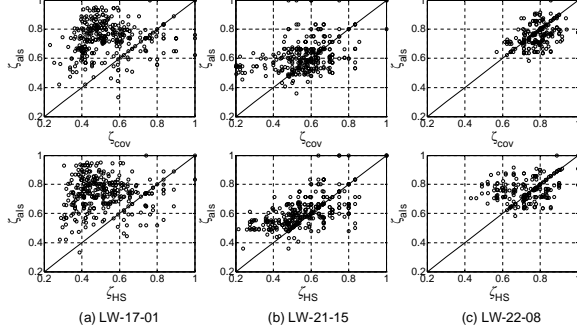


Fig. 7. Pairwise overlap comparison  $\zeta_{ALS}$  vs.  $\zeta_{COV}$  (top) and  $\zeta_{ALS}$  vs.  $\zeta_{HS}$  (bottom) for LW-17-01 (a), LW-21-15 (b) and LW-22-08 (c).

similarity between foreground and background. More visual comparisons are shown in Fig. 8(a). We see that  $AKF_{als}$  offers the best position and size estimation except in the AMCOM sequence LW-22-08, where  $AKF_{cov}$  is slightly better due to the lack of well defined structure in the histogram-based appearance, as shown in Fig. 6(e). We observe similar results against the SENSIAC data as shown in Fig. 8(b), where the  $AKF_{als}$  tracker outperforms the other two algorithms in most cases (except for 1927-0011), and can effectively adapt to varying poses and sizes for long sequences (200-500 frames). The performance of  $AKF_{als}$  slightly deteriorates against the SENSIAC sequence 1927-0011, especially towards the end, due to the presence of a dust cloud that greatly affects the appearance learning process due to occlusion of the target.

TABLE IV  
OVERLAP METRIC VALUES OF THE THREE TRACKING ALGORITHMS.

Sequences	HS	$AKF_{cov}$	$AKF_{als}$
LW-15-NS	0.669	0.707	<b>0.714</b>
LW-17-01	0.547	0.596	<b>0.720</b>
LW-21-15	0.601	0.578	<b>0.620</b>
LW-14-15	0.676	0.682	<b>0.708</b>
LW-22-08	0.751	<b>0.770</b>	0.758
LW-20-18	0.689	0.753	<b>0.758</b>
LW-18-17	<b>0.704</b>	0.702	0.703
LW-19-06	0.670	0.685	<b>0.713</b>
MW-14-10	<b>0.802</b>	0.797	0.799
LW-20-04	0.715	0.711	<b>0.720</b>
AMCOM Average	0.682	0.698	<b>0.721</b>
1927-0001	0.776	0.777	<b>0.799</b>
1927-0002	0.727	0.813	<b>0.845</b>
1927-0005	0.750	0.751	<b>0.787</b>
1927-0009	0.816	0.849	<b>0.855</b>
1927-0011	<b>0.858</b>	0.829	0.852
1925-0001	0.866	0.836	<b>0.872</b>
1925-0002	0.797	0.790	<b>0.824</b>
1925-0006	0.875	0.879	<b>0.886</b>
1925-0009	0.801	0.843	<b>0.860</b>
1925-0012	0.943	0.943	<b>0.946</b>
SENSIAC Average	0.821	0.831	<b>0.852</b>

### C. Tracking performance of covariance descriptor

We also tested the covariance descriptor for IR tracking. The covariance descriptor was found to be robust and effective for object tracking in optical images. It was first proposed in [62] for object detection with significant advantages than histogram-based appearance models, and extended to tracking by augmenting with manifold learning-based model update [64], [65]. In IR tracking, the covariance descriptor involves local intensity, stdev, gradient, orientation and Laplacian information of the target area. Like the other three histogram-based appearance learning algorithms, this descriptor was combined with the particle filter-based tracking algorithm [15]. Tracking results obtained using the covariance descriptor are shown in Fig. 9, where no learning was involved. We observe that the covariance tracker is able to maintain reasonable track lock against the target in LW-17-01, but fails to track the dim target in LW-15-NS. In both sequences, the tracker encountered difficulty in estimating the target size. The small target size, weak texture, and absence of color significantly reduced the effectiveness of the covariance descriptor for tracking small targets in IR imagery.

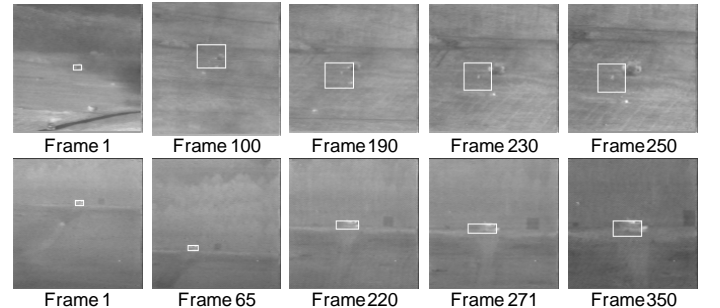


Fig. 9. Tracking results for two AMCOM sequences using the covariance tracker. Top: LW-15-NS and Bottom: LW-17-01.



Fig. 8. Tracking results against five AMCOM sequences (a) (from top to bottom: LW-15-NS, LW-17-01, LW-21-15, LW-14-15 and LW-22-08) and five SENSIAIC sequences (b) (1925-0009, 1927-0001, 1927-0002, 1927-0011 and 1925-0002). The top row of each image shows the observed frame and the bottom row depicts the track gates corresponding to the Ground truth (top-left), HS (top-right),  $AKF_{cov}$  (bottom-left),  $AKF_{als}$  (bottom-right).

#### D. Further Discussion

The HS method is usually encumbered by the drifting problem during incremental learning.  $AKF_{cov}$ , which assumes the same noise statistics for all histogram bins and estimates only the process noise without considering PSD conditions, results in a suboptimal Kalman gain. Its performance is marginally better than that of HS.  $AKF_{als}$ , which estimates both process and observation noises with PSD conditions for each histogram bin, is able to follow the modes and variations of the histogram during tracking and supports effective appearance learning. However, when the histogram lacks strong modes

and has widespread and small bin values, such as in the two AMCOM sequences LW-22-08 and MW-14-10, or when the histogram is not well-structured due to background clutter (as in SENSIAIC sequence 1927-0011), all three methods are comparable. This is mainly because the poor structure of the histogram evolution may invalidate the Kalman filter assumptions, while HS remains effective by incorporating the most recent observation for appearance learning when the histogram is poorly defined. This observation justifies the use of HS for learning the stdev histograms, which are normally characterized by weak structure.

TABLE V

MEAN ERROR OF THE STATE VARIABLES AVERAGED OVER THE LENGTH OF THE SEQUENCE FROM 50 MONTE CARLO RUNS USING THREE DIFFERENT ALGORITHMS FOR THE AMCOM (THE FIRST 10 SEQUENCES) AND SENSIAC (THE SECOND 10 SEQUENCES) DATASETS.

Algorithms	HS				AKF <sub>cov</sub>				AKF <sub>als</sub>			
Tracking errors	$x$	$y$	$s^x$	$s^y$	$x$	$y$	$s^x$	$s^y$	$x$	$y$	$s^x$	$s^y$
LW-15-NS	1.019	1.817	1.906	2.732	0.860	1.511	1.644	2.396	<b>0.801</b>	<b>1.461</b>	<b>1.423</b>	<b>2.339</b>
LW-17-01	2.406	3.415	2.104	<b>3.016</b>	2.145	3.005	2.101	3.163	<b>1.213</b>	<b>2.110</b>	<b>1.376</b>	3.033
LW-21-15	0.970	1.653	<b>2.624</b>	2.941	1.135	1.812	2.799	3.113	<b>0.893</b>	<b>1.300</b>	2.786	<b>2.575</b>
LW-14-15	<b>0.889</b>	0.815	3.160	2.137	0.932	<b>0.787</b>	2.981	2.157	1.099	0.801	<b>2.660</b>	<b>1.787</b>
LW-22-08	<b>1.167</b>	0.868	1.684	<b>2.049</b>	1.202	0.843	<b>1.070</b>	2.232	1.200	<b>0.839</b>	1.363	2.175
LW-20-18	3.230	1.831	1.657	1.953	0.901	1.095	<b>1.307</b>	1.766	<b>0.599</b>	<b>1.084</b>	1.439	<b>1.754</b>
LW-18-17	<b>1.269</b>	1.722	<b>0.733</b>	2.949	1.303	1.838	0.859	2.611	1.425	<b>1.679</b>	1.087	<b>2.252</b>
LW-19-06	1.977	1.545	1.566	1.544	0.797	0.764	1.681	1.454	<b>0.694</b>	<b>0.709</b>	<b>1.536</b>	<b>1.279</b>
MW-14-10	<b>0.628</b>	0.789	1.648	1.691	0.756	0.806	1.638	1.789	<b>0.778</b>	<b>1.629</b>	<b>1.267</b>	<b>1.607</b>
LW-20-04	0.702	0.954	<b>0.940</b>	1.528	0.697	0.937	1.071	1.614	<b>0.688</b>	<b>0.907</b>	1.006	<b>1.357</b>
AMCOM Average	1.426	1.541	1.802	2.254	1.073	1.340	1.715	2.230	<b>0.939</b>	<b>1.167</b>	<b>1.630</b>	<b>2.016</b>
1927-0001	0.504	1.162	1.965	5.830	0.629	0.902	1.915	5.258	<b>0.473</b>	<b>0.853</b>	<b>1.901</b>	<b>4.443</b>
1927-0002	1.835	2.276	0.001	6.024	0.897	2.147	0.000	<b>5.571</b>	<b>0.528</b>	<b>2.139</b>	<b>0.000</b>	5.580
1927-0005	2.335	<b>1.325</b>	0.000	4.728	2.188	2.092	0.000	4.864	<b>1.928</b>	1.622	<b>0.000</b>	<b>3.694</b>
1927-0009	1.686	1.556	0.089	2.721	1.421	1.317	<b>0.000</b>	1.979	<b>1.344</b>	<b>1.289</b>	0.003	<b>1.898</b>
1927-0011	<b>0.514</b>	2.384	0.004	4.969	1.065	<b>2.163</b>	0.018	4.484	0.702	2.425	<b>0.000</b>	<b>4.129</b>
1925-0001	0.848	2.116	0.000	<b>2.639</b>	1.204	1.890	0.000	3.669	<b>0.794</b>	<b>1.621</b>	<b>0.000</b>	3.733
1925-0002	0.859	3.057	1.750	3.958	<b>0.521</b>	3.780	1.750	3.197	0.526	<b>2.614</b>	<b>1.750</b>	<b>2.998</b>
1925-0006	<b>0.572</b>	3.359	0.000	<b>1.053</b>	0.700	2.629	0.000	1.589	0.616	<b>2.618</b>	<b>0.000</b>	1.449
1925-0009	1.376	5.177	0.039	<b>5.021</b>	1.337	2.355	0.000	5.427	<b>1.121</b>	<b>2.044</b>	<b>0.000</b>	5.210
1925-0012	<b>0.379</b>	<b>0.781</b>	0.070	2.459	0.385	0.834	0.070	2.312	0.385	0.822	<b>0.070</b>	<b>2.119</b>
SENSIAC Average	1.090	2.319	0.392	3.940	1.035	2.011	0.375	3.835	<b>0.842</b>	<b>1.805</b>	<b>0.372</b>	<b>3.525</b>

#### IV. CONCLUSION

We have presented a new IR target tracking algorithm that achieves state-of-the-art performance against extremely challenging infrared imagery. To the best of our knowledge, this is the first work reporting both near sub-pixel tracking accuracy and low size estimation error (1-2 pixels) against the challenging AMCOM IR closure sequences and the newly released SENSIAC MWIR sequences. The proposed approach encapsulates several recent innovations in target tracking as well as Kalman filtering in a joint tracking and learning framework. Specifically, the dual foreground-background target model is shown to be effective for enhancing the tracker sensitivity and robustness. Moreover, the new AKF<sub>als</sub> appearance learning method outperforms two existing histogram-based appearance learning techniques, viz., HS and AKF<sub>cov</sub>, as well as the recent covariance tracker that is often used against optical imagery.

#### REFERENCES

- [1] T. Han, M. Liu, and T. Huang, "A drifting-proof framework for tracking and online appearance learning," in *Proc. IEEE Workshop Applications Comput. Vision*, 2007.
- [2] J. Pan and B. Hu, "Robust object tracking against template drift," in *Proc. IEEE Int'l. Conf. Image Process.*, 2007, pp. 353–356.
- [3] L. Matthews, T. Ishikawa, and S. Baker, "The template update problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 810–815, 2004.
- [4] L. Latecki and R. Miezianko, "Object tracking with dynamic template update and occlusion detection," in *Proc. Int'l. Conf. Pattern Recog.*, Hong Kong, China, Aug. 20-24, 2006, vol. 1, pp. 556–560.
- [5] G. Hager and P. Belhumeur, "Real-time tracking of image regions with changes in geometry and illumination," in *Proc. IEEE Int'l. Conf. Comput. Vision, Pattern Recog.*, Jun. 18-20 1996, pp. 403–410.
- [6] Z. Peng, Q. Zhang, and A. Guan, "Extended target tracking using projection curves and matching pel count," *Optical Eng.*, vol. 46, no. 6, pp. 066 401–1 – 066 401–6, Jun. 2007.
- [7] C. Johnston, N. Mould, J. Havlicek, and G. Fan, "Dual domain auxiliary particle filter with integrated target signature update," in *Proc. IEEE Int'l. Conf. Comput. Vision, Pattern Recog. Workshops*, Jun. 20-25, 2009, pp. 54–59.
- [8] J. Khan and M. Alam, "Efficient target detection in cluttered FLIR imagery," in *Optical Pattern Recog. XVI*, ser. Proc. SPIE, D. Casasent and T.-H. Chao, Eds., vol. 5816, 2005, pp. 39–53.
- [9] S. Yi and L. Zhang, "A novel multiple tracking system for UAV platforms," in *ISR Systems and Applications III*, ser. Proc. SPIE, D. Henry, Ed., vol. 6209, 2006, 8 pp.
- [10] A. Dawoud, M. Alam, A. Bal, and C. Loo, "Decision fusion algorithm for target tracking in infrared imagery," *Optical Eng.*, vol. 44, pp. 026 401–1–8, Feb. 2005.
- [11] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, p. 13, 2006.
- [12] C. del Blanco, F. Jaureguizar, N. Garcia, and L. Salgado, "Robust automatic target tracking based on a Bayesian ego-motion compensation framework for airborne FLIR imagery," in *Polarimetric and Infrared Infrared Processing for ATR*, ser. Proc. SPIE, F. Sadjadi and A. Mahalanobis, Eds., vol. 7335, 2009, 12 pp.
- [13] N. Mould, C. Nguyen, C. Johnston, and J. Havlicek, "Online consistency checking for AM-FM target tracks," in *Proc. SPIE/IS&T Conf. Computational Imaging VI*, ser. Proc. SPIE, C. Bouman, E. Miller, and I. Pollak, Eds., vol. 6814, 2008, 12 pp.
- [14] V. Venkataraman, G. Fan, and X. Fan, "Target tracking with online feature selection in FLIR imagery," in *Proc. Int'l. Workshop Object Tracking, Class. Beyond the Visible Spectrum (in conjunction with CVPR2007)*, 2007.
- [15] V. Venkataraman, G. Fan, X. Fan, and J. P. Havlicek, "Appearance learning by adaptive Kalman filters for FLIR tracking," in *Proc. Int'l. Workshop Object Tracking, Class. Beyond the Visible Spectrum (in conjunction with CVPR2009)*, 2009.
- [16] A. Jepson, D. Fleet, and T. El-Maraghi, "Robust online appearance models for visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1296–1311, Oct. 2003.
- [17] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–577, 2003.
- [18] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms," in *Proc. IEEE Int'l. Conf. Comput. Vision, Pattern Recog.*, 1998, pp. 232–237.
- [19] J. Shaik and K. Iftekharuddin, "Automated tracking and classification of infrared images," in *Proc. IEEE Int'l. Joint Conf. Neural Netw.*, vol. 2, 2003, pp. 1201–1206.
- [20] A. Dawoud, M. Alam, A. Bal, and C. Loo, "Target tracking in infrared imagery using weighted composite reference function-based decision fusion," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 404–410, 2006.
- [21] Z. Wang, Y. Wu, J. Wang, and H. Lu, "Target tracking in infrared image sequences using diverse AdaBoostSVM," in *Proc. Int'l. Conf. Innovative Comput., Inf., Control*, 2006, pp. 233–236.

- [22] A. Yilmaz, X. Li, and M. Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1531–1536, Nov 2004.
- [23] A. Yilmaz, K. Shafique, and M. Shah, "Target tracking in airborne forward looking infrared imagery," *Image Vision Comput.*, vol. 21, no. 7, pp. 623–635, 2003.
- [24] P. S. Maybeck and S. K. Rogers, "Adaptive tracking of multiple hot-spot target IR images," *IEEE Trans. Autom. Control*, vol. AC-28, no. 10, pp. 937–943, Oct. 1983.
- [25] H. Nguyen and A. Smeulders, "Fast occluded object tracking by a robust appearance filter," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1099–1104, Aug 2004.
- [26] S. Lankton, J. Malcolm, A. Nakhmani, and A. Tannenbaum, "Tracking through changes in scale," in *Proc. IEEE Int'l. Conf. Image Process.*, 2008, pp. 241–244.
- [27] B. Han and L. Davis, "On-line density-based appearance modeling for object tracking," in *Proc. IEEE Int'l. Conf. Comput. Vision*, vol. 2, 2005, pp. 1492–1499.
- [28] R. Collins, "Mean-shift blob tracking through scale space," in *Proc. IEEE Int'l. Conf. Comput. Vision, Pattern Recog.*, vol. 2, 2003, pp. 234–240.
- [29] Z. Zivkovic and B. Krose, "An EM-like algorithm for color-histogram-based object tracking," in *Proc. IEEE Int'l. Conf. Comput. Vision, Pattern Recog.*, vol. 1, 2004, pp. 798–803.
- [30] N. Peng, J. Yang, and Z. Liu, "Mean shift blob tracking with kernel histogram filtering and hypothesis testing," *Pattern Recog. Lett.*, vol. 26, no. 5, pp. 605–614, Apr. 2005.
- [31] E. Maggio and A. Cavallaro, "Hybrid particle filter and mean shift tracker with adaptive transition model," in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Process.*, vol. 2, 2005, pp. 221–224.
- [32] C. Zhang and Y. Rui, "Robust visual tracking via pixel classification and integration," in *Proc. Int'l. Conf. Pattern Recog.*, vol. 3, 2006, pp. 37–42.
- [33] I. Leichter, M. Lindenbaum, and E. Rivlin, "Tracking by affine kernel transformations using color and boundary cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 164–171, Jan 2009.
- [34] J. Tu, H. Tao, and T. Huang, "Online updating appearance generative mixture model for meanshift tracking," *Mach. Vision Appl.*, vol. 20, no. 3, pp. 163–173, 2009.
- [35] S. Avidan, "Ensemble tracking," in *Proc. IEEE Int'l. Conf. Comput. Vision, Pattern Recog.*, vol. 2, 2005, pp. 494–501.
- [36] H. Chen, T. Liu, and C. Fuh, "Probabilistic tracking with adaptive feature selection," in *Proc. Int'l. Conf. Pattern Recog.*, vol. 2, 2004, pp. 736–739.
- [37] R. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1631–1643, Oct. 2005.
- [38] B. Han and L. Davis, "Robust observations for object tracking," in *Proc. IEEE Int'l. Conf. Image Process.*, vol. 2, Sep 2005, pp. 442–445.
- [39] H. T. Nguyen and A. W. Smeulders, "Robust tracking using foreground-background texture discrimination," *Int. J. Comput. Vision*, vol. 69, no. 3, pp. 277–293, 2006.
- [40] H. Nguyen, M. Worrington, and R. van der Boomgaard, "Occlusion robust adaptive template tracking," in *Proc. IEEE Int'l. Conf. Comput. Vision*, vol. 1, Jul. 07-14, 2001, pp. 678–683.
- [41] R. Mehra, "Approaches to adaptive filtering," *IEEE Trans. Autom. Control*, vol. 17, no. 5, pp. 693–698, Oct. 1972.
- [42] B. Odelson, R. Rajamani, and B. Rawlings, "A new autocovariance least-squares method for estimating noise covariances," *Automatica*, vol. 42, no. 2, pp. 303–308, 2006.
- [43] B. Odelson, A. Lutz, and J. Rawlings, "The autocovariance least-squares method for estimating covariances: application to model-based control of chemical reactors," *IEEE Trans. Control Syst. Technol.*, vol. 14, no. 3, pp. 532–540, 2006.
- [44] M. Swain and D. Ballard, "Indexing via color histograms," in *Proc. IEEE Int'l. Conf. Comput. Vision*, 1990, pp. 390–393.
- [45] G. Noriega and S. Pasupathy, "Adaptive estimation of noise covariance matrices in real-time preprocessing of geophysical data," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 5, pp. 1146–1159, Sep 1997.
- [46] X. Li and Y. Bar-Shalom, "A recursive multiple model approach to noise identification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 671–684, Jul 1994.
- [47] A. H. Mohamed and K. P. Schwarz, "Adaptive Kalman filtering for INS/GPS," *J. Geodesy*, vol. 73, pp. 193–203, 1999.
- [48] H. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer, 1994.
- [49] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Autom. Control*, vol. 15, no. 2, pp. 175–184, Apr. 1970.
- [50] B. Carew and P. Bélanger, "Identification of optimum filter steady-state gain for systems with unknown noise covariances," *IEEE Trans. Autom. Control*, vol. 18, no. 6, pp. 582–587, Dec. 1973.
- [51] C. Neethling and P. Young, "Comments on 'identification of optimum filter steady-state gain for systems with unknown noise covariances'," *IEEE Trans. Autom. Control*, vol. 19, no. 5, pp. 623–625, Oct. 1974.
- [52] B. Odelson, "Estimating disturbance covariances from data for improved control performance," Ph.D. dissertation, University of Wisconsin, 2003.
- [53] M. Rajamani, J. Rawlings, and T. Soderstrom, "Application of a new data-based covariance estimation technique to a nonlinear industrial blending drum," Texas-Wisconsin Modeling and Control Consortium, Tech. Report 2007-03, Sep. 2007.
- [54] D. L. Donoho, S. Mallat, R. von Sachs, and Y. Samuelides, "Locally stationary covariance and signal estimation with macrotils," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 614–627, Mar. 2003.
- [55] I. Tabus and A. Vasiache, "Low bit rate vector quantization of outlier contaminated data based on shells of Golay codecs," in *Proc. Data Compression Conf.*, Mar. 16-18, 2009, pp. 302–311.
- [56] T. V. Ramabadran and D. Sinha, "Speech data compression through sparse coding of innovations," *IEEE Trans. Speech, Audio Process.*, vol. 2, no. 2, pp. 274–284, Apr. 1994.
- [57] S. Boucheron and M. R. Salamatian, "About priority encoding transmission," *IEEE Trans. Info. Theory*, vol. 46, no. 2, pp. 699–705, Mar. 2000.
- [58] M. Raginsky, "Joint fixed-rate universal lossy coding and identification of continuous-alphabet memoryless sources," *IEEE Trans. Info. Theory*, vol. 54, no. 7, pp. 3059–3077, Jul. 2008.
- [59] A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard, "On the importance of combining wavelet-based nonlinear approximation with coding strategies," *IEEE Trans. Info. Theory*, vol. 48, no. 7, pp. 1895–1921, Jul. 2002.
- [60] G. I. Shamir and D. J. Costello, "Asymptotically optimal low-complexity sequential lossless coding for piecewise-stationary memoryless sources – Part I: the regular case," *IEEE Trans. Info. Theory*, vol. 46, no. 7, pp. 2444–2467, Jul. 2000.
- [61] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. IEEE Int'l. Conf. Comput. Vision, Pattern Recog.*, vol. 2, 2000, pp. 142–149.
- [62] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Proc. 9th European Conf. Comput. Vision*, 2006, pp. 589–600.
- [63] K. She, G. Bebis, H. Gu, and R. Miller, "Vehicle tracking using on-line fusion of color and shape features," in *Proc. IEEE Int'l. Conf. Intell. Transp. Syst.*, 2004, pp. 731–736.
- [64] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on lie algebra," in *Proc. IEEE Int'l. Conf. Comput. Vision, Pattern Recog.*, vol. 1, 2006, pp. 728–735.
- [65] X. Li, W. Hu, Z. Zhang, X. Zhang, M. Zhu, and J. Cheng, "Visual tracking via incremental log-Euclidean Riemannian subspace learning," in *Proc. IEEE Int'l. Conf. Comput. Vision, Pattern Recog.*, 2008.

PLACE  
PHOTO  
HERE

**Vijay Venkataraman** received the B. Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, in 2003 and the M.S and Ph.D. degrees in electrical engineering from Oklahoma State University, Stillwater, OK, in 2005 and 2010, respectively. Currently he is a research staff member at Utopiacompression Corporation, Los Angeles, CA. His research interests include computer vision, machine learning and vision systems for unmanned platforms.

PLACE  
PHOTO  
HERE

**Guoliang Fan (S'97, M'01, SM'05)** received the B.S. degree in Automation Engineering from Xi'an University of Technology, Xi'an, China, M.S. degree in Computer Engineering from Xidian University, Xi'an, China, and Ph.D. degree in Electrical Engineering from University of Delaware, Newark, DE, USA, in 1993, 1996, and 2001, respectively.

Since 2001, Dr. Fan has been with the School of Electrical and Computer Engineering at Oklahoma State University (OSU), Stillwater, OK, where currently he is a Professor and the Cal & Marilyn

Vogt Professor of Engineering. Dr. Fan is directing the Visual Computing and Image Processing Laboratory (VCIPL). His research interests include image processing, machine learning, computer vision, biomedical imaging and remote sensing applications.

Dr. Fan is a recipient of the 2004 National Science Foundation (NSF) CAREER award. He received the Halliburton Excellent Young Teacher Award in 2004, the Halliburton Outstanding Young Faculty Award in 2006 from the College of Engineering at OSU, and the Outstanding Professor Award from IEEE-OSU in 2008 and 2011. He is an associate editor of IEEE Trans. Information Technology in Biomedicine, EURASIP Journal on Image and Video Processing and ISRN Machine Vision.

PLACE  
PHOTO  
HERE

**Joseph P. Havlicek (S'84, M'88, SM'98)** received the B.S. and M.S. degrees from Virginia Tech, Blacksburg, and the Ph.D. degree from the University of Texas, Austin.

He has been a software developer with Management Systems Laboratories, Blacksburg, VA, and IBM, Austin, TX. From 1987 to 1997 he was with the Naval Research Laboratory, Washington, DC, where he was a recipient of the Department of the Navy Award of Merit for Group Achievement for work in infrared threat warning sensors. Since 1997

he has been with the School of Electrical and Computer Engineering at the University of Oklahoma where he is currently the Williams Presidential Professor. His research interests include signal and image processing and intelligent transportation systems.

Dr. Havlicek served as a Technical Area Chair for the 2012 IEEE International Conference on Image Processing and the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing. He is currently an Associate Editor of IEEE Transactions on Image Processing and IEEE Transactions on Industrial Informatics.

PLACE  
PHOTO  
HERE

**Xin Fan (M'06)** was born in 1977. He received the B.E. and Ph.D. degrees in Information and Communication Engineering from Xi'an Jiaotong University, China, in 1998 and 2004 respectively.

From May 2006 to Dec. 2007, he worked in Oklahoma State University as a postdoctoral research fellow. In Jan. 2008, he moved to the University of Texas Southwestern Medical Center in Dallas for his second postdoctoral training. He joined the School of Software at Dalian University of China in November 2001 where he is an associate professor and the head

of department of Digital Media Technology. His research interests include Bayesian estimation and machine learning, and their applications to target tracking, image super-resolution, and DTI-MR image analysis.

PLACE  
PHOTO  
HERE

**Yan Zhai (S'00, M'07, SM'10)** received the B.S. degree from Tsinghua University, Beijing, China, the M.S. degree in mechanical engineering from Oklahoma State University, Stillwater, and the Ph.D. degree in electrical and computer engineering from the University of Oklahoma, Norman, in 1998, 2002, and 2007, respectively.

Dr. Zhai is currently with Micron Technology as an Algorithm Development Specialist. His research interests include statistical signal processing, image processing, and computer vision. He has written a variety of research papers and has served as a reviewer for several national and international conferences and journals. Dr. Zhai received the 2011 Outstanding Young Engineer Award of the IEEE Instrumentation and Measurement Society. He is also a member of Sigma Xi and Tau Beta Pi.

PLACE  
PHOTO  
HERE

**Mark B. Yeary (S'95, M'00, SM'03)** received the B.S. (honors), M.S., and Ph.D. degrees from Texas A&M University, College Station, TX in 1992, 1994, and 1999, respectively.

He is now with the University of Oklahoma's School of Electrical and Computer Engineering, where he has been recently promoted to the rank of full professor. He also holds the endowed Hudson-Torchmark Presidential Professorship. While with OU, he has served as a Principle Investigator (PI) or Co-PI on federal projects sponsored by the NSF, NASA, NOAA, AFOSR, ONR, etc. He specializes in the areas of digital signal processing (DSP), embedded DSP systems, and radar engineering.

He has spent ten summers, 2002 through 2011, with Raytheon near Dallas, TX as faculty researcher, which includes target detection and tracking studies. Dr. Yeary will join the Massachusetts Institute of Technology's (MIT's) Lincoln Laboratory in the fall of 2012 and spring of 2013 semesters as a visiting research scientist. Dr. Yeary is a member of the Tau Beta Pi and Eta Kappa Nu honor societies and is a licensed professional engineer (PE). By invitation, he was selected to participate in the U.S. National Academy of Engineering's Foundations of Engineering Education Symposium in 2010.