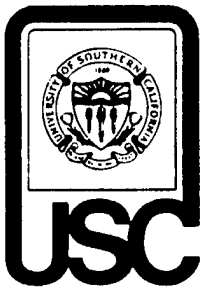
TECHNICAL REPORT CAMS–98.9.1

Center for Applied Mathematical Sciences



Denney Research Building 308
University of Southern California
Los Angeles, CA 90089-1113
Telephone: (213) 740-2395

DETECTION ALGORITHMS AND TRACK BEFORE DETECT ARCHITECTURE BASED ON NONLINEAR FILTERING FOR INFRARED SEARCH AND TRACK SYSTEMS

TECHNICAL REPORT CAMS–98.9.1

SKIRMANTAS KLIGYS, BORIS ROZOVSKY, AND ALEXANDER TARTAKOVSKY

CENTER FOR APPLIED MATHEMATICAL SCIENCES
University of Southern California
Los Angeles, CA 90089–1113

SEPTEMBER 1998

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	9-1-98	Technical

4. TITLE AND SUBTITLE
Detection Algorithms and Track Before Detect Architecture Based on Nonlinear Filtering For Infrared Search and Track Systems

5. FUNDING NUMBERS
DAAG55-98-1-0418

6. AUTHOR(S)
Skirmantas Kligys, Boris Rozovsky, and Alexander Tartakovsky

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Center for Applied Mathemaitcal Sciences
University of Southern California
Los Angeles, CA 90089-1113

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 38076.1-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

In this report we describe the developed computationally efficient algorithms and adaptive architecture with optimized overall performance (statistical and computational) for real-time reliable detection and tracking of low-observable targets in IRST systems. Despite the fact that we focus on an IRST against cruise missiles over land and sea cluttered backgrounds, the results are equally applicable to other sensors (e.g., Radar, Lidar) and other kind of targets (e.g. ballistic missiles). We concentrated on the three interrelated problems: (1) efficient clutter suppression; (2) development of the adaptive track-before-detect architecture based on optimal nonlinear filtering; (3) development of efficient algorithms for detection of a priori unknown number of targets that appear and disappear at unknown points in time. The detection algorithms are adaptive and use the estimates of the target location. These estimates are the results of target tracking (before detection) by the optimal spatio-temporal nonlinear filters. The corresponding nonlinear filtering algorithm is based on the spectral separation scheme and allows for the real time computation of the whole joint posterior distribution of the target location. The algorithms are tested for real heavy cluttered IR background with artificially inserted dim maneuvering targets. The developed algorithms show high performance even in very low SNR situations (up to 7dB after preprocessing and clutter removal).

14. SUBJECT TERMS
IRST, Sequential Detection, Track-Before-Detect, Nonlinear Filtering

15. NUMBER IF PAGES
52

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

CONTENTS

1. **Introduction** 2
2. **Problem Formulation and Background** 3
2.1. TbD Methods: Merits and Drawbacks 4
3. **Signal and Observation Modeling** 6
4. **Clutter Suppression** 7
4.1. Nonparametric Method 8
4.2. Semiparametric Filtering 8
5. **Optimal Nonlinear Filtering for TbD** 9
5.1. Modeling for TbD 9
5.2. Optimal Spatio-Temporal Nonlinear Filtering: The Basic Algorithm 10
5.3. Multi-Dimensional Spatio-Temporal Matched Filter as a Special Case of Nonlinear Filter 12
6. **Track Appearance/Disappearance Detection** 14
6.1. Preliminaries 14
6.2. Problem Formulation and Model Description 16
6.3. Detection Algorithm 1: Fixed Sliding Window 17
6.4. Detection Algorithm 2: Fully Sequential Procedure 19
6.5. Algorithm 3: Joint Detection of Track Appearance and Disappearance 21
6.6. Adaptive Detection Algorithms 23
6.7. Choice of Thresholds and Performance Evaluation 24
6.8. TbD Model with Spatio-Temporal Matched Filter 29
6.9. TbD Model with Optimal Nonlinear Filter 31
7. **Testing of the Developed Algorithms for IRST Data. Results of Experiments and Simulation** 32
7.1. Clutter Removal: Real IRST Data 32
7.2. Example 1: TbD of a Target Based on IRST Data 33
7.3. Example 2: TbD of a Surface Skimming Missile 41
7.4. Appearance/Disappearance Detection of a Skimming Missile 47
8. **Conclusion. Future Research** 50
9. **Acknowledgement** 50
References 51

1. Introduction

Cruise missiles over land and sea cluttered background are serious threats to Infrared Search and Track Systems (IRST's). In general, these threats are stealth in both the infrared and radio frequency bands. That is, their thermal infrared signature and their radar cross section can be quite small. Future predicted threats, i.e. the next generation of cruise missiles, will be even more difficult to detect at a sufficient range to counter. Further, low elevation trajectory objects, such as sea skimming missiles, have radar signals with large amounts of temporally and spatially correlated interference called multipath. This multipath problem remains an enormous obstacle to existing trackers. Hence, new technology is needed which will allow for the timely detection, tracking, and identification of such threats.

IRST systems are one component of a multisensor suite which can meet the technical challenge of the timely detection/track/identification of low signal-to-(noise+clutter) ratio (S(N+C)R) targets. The multisensor suite should include an IRST, Radar, and a coherent laser (Lidar). We envision a cueing hierarchy where the IRST can cue the Radar or the Radar cues the IRST. Once a candidate track is established the Lidar can be used to identify the target by its micro doppler signature.

In this report we describe the developed computationally efficient algorithms and adaptive architecture with optimized overall performance (statistical and computational) for real-time reliable detection and tracking of low-observable targets in IRST systems. Despite the fact that we focus on an IRST against cruise missiles over land and sea cluttered backgrounds, the results are equally applicable to other sensors (e.g., Radar, Lidar).

In the research we concentrated on the three interrelated problems: (1) efficient clutter suppression; (2) development of the adaptive track-before-detect (TbD) architecture based on optimal nonlinear filtering (ONF); (2) development of efficient algorithms for detection of *a priori* unknown number of targets that may appear and disappear at unknown points in time.

The report is organized as follows. In Section 2 we formulate the problem, describe a structure of the system to be developed, outline popular track-before-detect methods that are in current use and suggest an alternative method, which is based on the optimal nonlinear filtering. In Section 3 we describe basic models and assumptions on signals and clutter that are used in developing of signal processing algorithms (clutter suppression, track-before-detect) and detection algorithms. In Section 4, two clutter removal algorithms are presented based on nonparametric and semi-parametric spatio-temporal filtering. Section 5 is especially important. Here we describe an optimal nonlinear filtering technique that is used for track-before-detect of very low observable targets. Also, we show that the proposed method is a complete generalization of the multi-dimensional (spatio-temporal) matched filtering (particularly, 3D matched filter). Furthermore it is shown that the spatio-temporal matched filter coincides with the developed optimal nonlinear filter when a target moves according to deterministic trajectory. In Section 6 we develop three track appearance/disappearance algorithms all of which take into account requirements specific for surveillance systems. The results of simulation and processing real IR data obtained from

SPAWAR (Space & Naval Warfare Systems Center, San Diego, CA) are presented in Section 7. Finally in Section 8 we provide a conclusion and the plan of our future research.

2. Problem Formulation and Background

The generalized block-diagram of the system under investigation is shown in Figure 1. We develop both the signal processing architecture (clutter removal algorithms and TbD algorithms) and track detection algorithms.

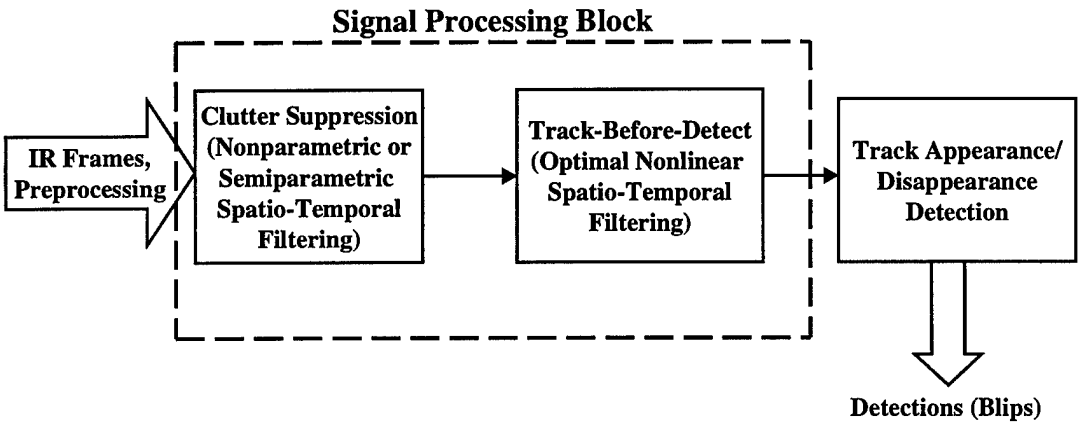


Figure 1. Generalized block-diagram of the developed system

New algorithms concerning the stages of data processing specified above are developed under the following realistic conditions.

- Cluttered background is much more intensive than both equivalent intrinsic (instrumental) noise of the sensor and signal intensity of the targets to be detected. This causes a necessity of practically complete suppression of a clutter.
- Exterior conditions of observation are characterized by an extremely high variability and prior uncertainty and may not be predicted with sufficient accuracy.
- Prior information that is needed to develop ideal (Bayes) data processing algorithms is not available. Particularly, statistical models of signals and moreover exterior background as well as the models of changing target situation are extremely unreliable. Such models can be useful only as tools for performance evaluation in certain scenarios but not for development of data processing algorithms. Practical data processing algorithms should be developed on the base of sequential application of robust, adaptive and minimax methods that are invariant to the prior uncertainty.
- In practice the estimated parameters of targets (for example, trajectory parameters) should be guaranteed for any degree of prior uncertainty. Specifically, each estimate should be supplemented with an appropriate domain of minimum size that contains the real unknown value of estimated parameter with 100% (or at least $(1-\varepsilon)\%$) assurance.

2.1. **TbD Methods: Merits and Drawbacks.** The most challenging problem for an IRST system is the detection of a maneuvering target against a strong clutter background. To illustrate the importance of this task, we remark that under certain conditions *a few seconds decrease* in the time it takes to detect a sea/surface skimming cruise missile can yield a significant increase in the *probability of raid annihilation.* The problem of detection is extremely difficult in low S(N+C)R when localization of the target based upon a single non-stationary image is impossible. In this case one has to align successive frames according to typical patterns of target dynamics and any results of "preliminary" tracking. This approach to detection of a low S(N+C)R target is usually referred to as "track-before-detect" (TbD). Its success depends crucially on the quality of the "preliminary" tracker. Thus, the development of the efficient coherent signal processing based on the TbD methodology becomes crucial point in the low-observable target detection problem. In contrast to other TbD methods, we solve this problem by applying optimal nonlinear filtering approach.

We now overview several popular methods for tracking before detection that are in current use.

2.1.1. ***Spatio-Temporal Matched Filters and Velocity Filter Banks.*** Given a sequence of frames, consider the problem of detecting a small (unresolved) target against a clutter background. The probabilities of errors can be decreased by applying a 3−D (spatio-temporal) matched filter prior to detection (thresholding) operation. If the spatial distribution of the target and its velocity remain unchanged (i.e., if targets move with known constant speed along a line on the plane) and the noise and the clutter are Gaussian processes, the 3−D filter is the optimal method of detection, see [38]. It is easily shown that the same result is true for more dimensions, i.e. a $(m+1)$−D matched filter is an optimal method of processing under aforementioned conditions (m is the spatial dimensionality). Also, under these conditions the "target" component of the multi-dimensional matched filter separates into spatial and temporal components. The temporal component is usually called a velocity filter.

Typically the target velocity is unknown and hence the single velocity filter cannot be used. This problem is usually overcome by hypothesizing velocities and implementing a velocity filter bank (see, e.g., [43, 48, 50]).

One of the main drawbacks of spatio-temporal matched filters and other modifications such as banks of assumed velocity filters is that they are not able to work with maneuvering targets. Performance of the algorithms is substantially degraded in the presence of velocity mismatch or in event of target maneuver.

2.1.2. ***Dynamic Programming Methods.*** The Dynamic Programming methods for TbD showed a big advantage over the conventional MHT method and over the 3−D matched filter with velocity mismatch [1, 2, 7, 21, 57]. Particularly the results of Fernandez et al. [21] show that application of the Viterbi TbD algorithm over 10 frames of IR data yields about a 7 dB improvement in detection sensitivity over conventional thresholding/peak-detection procedures.

This approach avoided problems with velocity mismatch and could handle targets with slow maneuvers. However, good performance of dynamic programming algorithms is observed for moderate S(N+C)R (over 3 dB after preprocessing and clutter suppression) with rapid degradation as S(N+C)R reduces further [57]. In addition, the computational complexity of these sophisticated methods is fairly high.

2.1.3. ***Extended Kalman Filter.*** To date, the extended Kalman filter (EKF) along with minor variations have been the dominant algorithm technology in real-time tracking. It is the basis for current practical single- or multi-target trackers for point equivalent targets. A major reason for its success has been the fact that the EKF has offered a reasonable compromise between real-time operation and accurate performance in many nonlinear problems. On the other hand, the EKF is a suboptimal and completely heuristic algorithm whose efficiency varies from case to case. For instance, the EKF is unstable in situations that involve acute maneuvering, missing measurement, low SNR, multipath, and many other situations where the posterior distribution may not be approximated well enough by a Gaussian distribution. It is very difficult, if even possible, to develop rigorous evaluation metrics for assessing the quality of data processing based on the EKF technology. Improvements to the EKF (e.g. iterated EKF) work satisfactory in a number of important applications where the EKF fails, but still, it is difficult to overcome the fundamental limitations of the EKF algorithm that stem from its dependence on the assumption that the posterior distribution may be well approximated by a Gaussian distribution. The typical posterior density built upon the realistic IR image is shown in Figure 17, Section 7 . One sees that it has multi-peak form and hence is very far from being Gaussian. Our experiments show that EKF is typically fails for this kind of data.

2.1.4. ***An Alternative TbD Method – Optimal Nonlinear Filtering.*** In spite of the aforementioned shortcomings, the above outlined methods remain the basis for the great majority of existing signal processing systems. In particular, until recently no other information technology was able to effectively compete with the EKF in target tracking.

However the situation has changed with recent advances in the mathematical theory and algorithmic support for *optimal nonlinear filtering* (ONF). These advances coupled with improvements to modern digital hardware technology make optimal nonlinear filtering an attractive alternative to the multi-dimensional matched filter, dynamic programming based algorithms and EKF in many practical and important applications. These include signal processing for infrared and acoustic sensors, imaging radar and sonar (including SAR and SAS), and other passive and active sensors [17, 45, 49].

Advanced optimal nonlinear filters can now provide:

- real-time operation;
- optimal, theoretically sound solutions of the full nonlinear problem;
- distributional versatility (no constraints on the form of prior or posterior probability distributions);

- superior accuracy and robustness;
- facility to effectively incorporate realistic physical models;
- explicit quantitative performance metrics (exact error estimates, confidence areas, etc.).

Our analysis of advanced algorithms based on ONF technologies shows great promise in: track-before-detect of unresolved targets in low S(N+C)R (up to −6dB after preprocessing); fusion of imaging and kinematic data for target identification; tracking of agile extended targets in cluttered environment as well as with certain other applications.

We argue that just as the EKF superseded the $\alpha-\beta-\gamma$ trackers, so the optimal nonlinear filter is set to replace the EKF as the dominant tracking technology within 10-15 years.

In Section 5 we describe the ONF techniques that will be used in defining and developing the appropriate technology for application in an end-to-end IRST signal processing architecture. The architecture will be sufficiently flexible to allow for an adaptive optimization of the signal/discrimination processors utilizing existing engineering parameters. This adaptive optimization will use the output of an EO/IR sensor diagnostic tool to utilize current meteorological and environmental information as well as current intelligence on likely threat scenarios.

3. Signal and Observation Modeling

It is assumed that a sensor has $m-$component resolution capability (for radar typically $m=6$, for IR/EO $m=4$). By $\boldsymbol{r}=(r_1,\dots,r_m)$ will be denoted a phase coordinate vector (for IRST typically angles and angle velocities of an object in a certain coordinate system). The sensor carries out a periodic surveillance of definite area in $m-$dimensional domain $\mathbb{D}^m\subset\mathbb{R}^m$, where $\mathbb{R}^m$ is the $m-$dimensional Euclidean space. After standard preprocessing the result of one observation step is a frame of measurements,

$$\begin{aligned}&\boldsymbol{Z}(n)=\boldsymbol{S}(n)+\boldsymbol{b}(n)+\boldsymbol{\eta}(n)\qquad n=1,2,\dots,\quad\text{or}\\&\boldsymbol{Z}(n)=\|Z_i(n)\|=\|S_i(n)+b_i(n)+\eta_i(n)\|\,,\quad i=1,\dots,N,\end{aligned}\tag{3.1}$$

where $\boldsymbol{S}(n)=\|S_i(n)\|$ is a signal from target, $\boldsymbol{b}(n)=\|b_i(n)\|$ is an exterior background (clutter), and $\boldsymbol{\eta}=\|\eta_i(n)\|$ is a noise of the sensor. (Here we assume that after preprocessing sampling of data is done in discrete points d_i, $i=1,\dots,N$, uniformly in the area $\mathbb{D}^m$ (i is a pixel).)

The noise is assumed to be zero mean and uncorrelated in both time n and space i, $\boldsymbol{E}\eta_i(n)=0$, $\boldsymbol{E}\eta_i(n)=\sigma_\eta^2$ ($\boldsymbol{E}$ is a symbol of expectation). The clutter is defined as

$$b_i(n)=b(\boldsymbol{r}_i+\boldsymbol{\delta}(n),n)\tag{3.2}$$

where $b(\boldsymbol{r},n)$ is a function describing the background (spatial distribution of the clutter) after preprocessing in the point $\boldsymbol{r}$, and $\boldsymbol{\delta}(n)=(\delta_1(n),\dots,\delta_m(n))$ is an unknown current bias of sensor coordinate system with respect to the reference one (due to the jitter).

The signal component is modeled as

$$S_i(n) = \sum_{j=1}^{k(n)} A_j(n) h(\boldsymbol{r}_i + \boldsymbol{\delta}(n) - \boldsymbol{r}_j(n)) \tag{3.3}$$

where $h(\boldsymbol{r})$ is a normalized sensor function; $k(n)$ is an unknown total number of targets at the moment n; $A_j(n)$ and $\boldsymbol{r}_j(n)$ are unknown signal intensity and coordinates of the jth target, respectively.

It is assumed that the clutter $b(\boldsymbol{r}, n)$ has a relatively big spatial variance (the change of $b(\boldsymbol{r}, n)$ between two nearest values of $\boldsymbol{r}_i$ is comparable with the maximum value of $b(\boldsymbol{r}, n)$). Besides, $b(\boldsymbol{r}, n)$ varies locally as fast as the signal function does (in spatial coordinates). Even in cases where these assumptions do not exactly hold, the algorithms that rely on them will be robust, which is the most important requirement. The second assumption shows that if only spatial information is used, then the background may be interpreted as a target. Thus spatial filtering alone is not sufficient for clutter suppression and temporal filtering is needed. We always assume that $b(\cdot)$ is an arbitrary unknown function of $\boldsymbol{r}$ and slowly changing function in n: there exists such T that $|b(\boldsymbol{r}, n+T) - b(\boldsymbol{r}, n)| \leq \sigma_\eta$. The latter assumption, which often holds, shows that variations of the clutter in time are caused mainly by uncontrolled vibrations $\boldsymbol{\delta}(n)$ of the sensor. These vibrations are unknown and unpredictable except for a natural restriction, $|\boldsymbol{\delta}(n)| \leq \Delta$, imposed on the absolute value of bias of coordinate system. No assumptions on statistical behavior of clutter is made. It is our belief that such popular models as homogeneous random field, especially Gaussian, are valuable only for purely academic research and lead to highly non-robust filtering algorithms.

Perhaps the most important feature of our approach to algorithm design is that we refuse to use any artificial and unreliable statistical models of $b(\boldsymbol{r}, n)$, $\boldsymbol{\delta}(n)$, $k(n)$, and $A_j(n)$. The algorithms developed on the basis of such models fail even for small deviation of a model from reality. The essence of the new suggested approach is the development of algorithms that are invariant and/or adaptive with respect to prior uncertainty. The specific feature of the problem is its extremely high dimensionality. The value of N can be of the order of $10^6 - 10^8$ and the total number of targets can be several hundreds. Thus, along with the mathematical issue of data processing, serious attention should be paid to the computational complexity, parallelization and HPC realization.

4. Clutter Suppression

Two classes of algorithms for background filtering are proposed. In either case there are two basic problems to be solved: (a) to transform the sequence of input frames into the new frame such that the clutter would be reduced and the signal would be preserved (clutter removal); (b) for every n the location of the coordinate system of the sensor (i.e. the bias $\boldsymbol{\delta}(n)$) should be estimated with maximum possible accuracy (jitter compensation). Below we propose two algorithms that allow the first problem to be solved effectively. Jitter compensation algorithms will be developed in the near future.

4.1. **Nonparametric Method.** The first class is relied on the nonparametric regression approach to estimation of the function $b(\boldsymbol{r}, n)$. Our goal is to build a nonparametric clutter estimate such that the residuals between the original data and its smoothed version (estimate) would be reasonably approximated by signal plus noise models. That is, the estimate $\hat{b}(\boldsymbol{r}, n)$ of the $b(\boldsymbol{r}, n)$ should be built in such a way that in the filtered frame $\tilde{Z}(\boldsymbol{r}, n) = Z(\boldsymbol{r}, n) - \hat{b}(\boldsymbol{r}, n)$ the signal would be preserved, while the clutter would be removed almost completely.

Kernel methods provide a powerful tool for such analysis due to both theoretic optimality (see [19, 23]) and computational transparency. In addition these methods are invariant to statistical properties and variations of clutter. Kernel estimators are weighted moving averages of observations

$$\hat{b}(\boldsymbol{r}, n) = \frac{1}{\prod_{i=1}^{m} N_i} \sum_{r_1, \ldots, r_m} Z(\boldsymbol{r}, n) K\left(\tfrac{l_1 - r_1}{N_1}, \ldots, \tfrac{l_N - r_2}{N_m}\right),$$

where N_i are window sizes in corresponding directions, $K(\cdot)$ is a deterministic $m-$dimensional kernel (recall that $\boldsymbol{r} = (r_1, \ldots, r_m)$). A product kernel with univariate kernels K_i, $i = 1, \ldots, m$, provides a popular example.

Note that in the 2D case ($m = 2$) a so called 'local mean removal' procedure used for clutter rejection in a series of papers by Reed et al. [38, 39] is equivalent to a kernel estimator with a product kernel K generated by uniform kernels $K_1(x) = K_2(x) = 0.5 I_{\{|x| \leq 1\}}$. It is shown in Section 4.1 that the proposed adaptive nonparametric methods based on recent advances in nonparametric estimation (see e.g. [18], [27]) perform much better than the aforementioned 'local mean removal' method.

4.2. **Semiparametric Filtering.** The second method (semiparametric) is based on the adaptive spatio-temporal auto-regression

$$\tilde{Z}_i(n) = Z_i(n) - \sum_{\tau=n-T}^{n} \sum_{l=0}^{L} a_l(\tau) Z_{il}(n) \tag{4.1}$$

where $\tilde{\boldsymbol{Z}}(n) = ||\tilde{Z}_i(n)||$ is the filtered frame of input data $\boldsymbol{Z}(n) = ||Z_i(n)||$; $Z_{il}(n)$ are the values of $Z_i(n)$ in some vicinity (spatial filtering window) of points $\boldsymbol{r}_i$; L is the number of points $\boldsymbol{r}_j$ in the window; $a_l(\tau)$ are some coefficients.

The major problem is the choice of optimal coefficients $a_l(\tau)$. They must be calculated adaptively to guarantee minimum of empirical mean-square (MS) value of the filtering residual noise (internal noise and background residual) for every time moment n. Such a criterion provides minimum of MS residual noise in the frame for every time n and is invariant with respect to statistical properties of the background $b(\boldsymbol{r}_i + \boldsymbol{\delta}(n), n)$ and noise. The algorithm consists of:

1. Computation of empirical correlation matrix ($TL \times TL$) for spatial window L and temporal window T.

2. Calculation of the optimal weights $a_l(\tau)$.

3. Linear transform of $Z_i(n)$ into $\tilde{Z}_i(n)$ according to (4.1).

In general the algorithm is nonlinear because of the nonlinear dependence of $a_l(\tau)$ on Z_i. This algorithm allows us to suppress any background regardless of its variation in time. However, it solves only the first part of the problem – it does not estimate a drift of sensor coordinate system to correct coordinate measurements on the stage of detection (jitter compensation). An algorithm that allows us to estimate $\boldsymbol{\delta}(n)$ and hence to compensate jitter will be developed in the near future.

5. Optimal Nonlinear Filtering for TbD

5.1. **Modeling for TbD.** Our philosophy in addressing the low S(N+C)R detection problem (regardless of the type of a sensor being used) is that the farther the thresholding operation may be postponed the better. In the context of track before detect, we assume that the data are collected at discrete times while the target dynamics is a continuous time process. To be specific, assume that the measurements $Z(k,\boldsymbol{r}) = Z(t_k,\boldsymbol{r})$ are collected at discrete time moments t_k, $k = 0,1,2,\ldots$ and the relationship between the observation and target location is modeled by a nonlinear measurement equation of the form

$$Z(k,\boldsymbol{r}) = S(k,\boldsymbol{X}_k,\boldsymbol{r}) + b(k,\boldsymbol{r}) + V(k,\boldsymbol{r})\,, \tag{5.1}$$

where, as before, $\boldsymbol{r}$ represents the spatial coordinate in the phase space, $b(k,\boldsymbol{r})$ is a sequence of deterministic (unknown) functions representing cluttered background, by $\boldsymbol{X}_k$ is denoted the true location of the target at time $t = t_k$ and by $V(k,\boldsymbol{r})$ is denoted the measurement noise process (sensor noise). For simplicity we first suppose that there may be only one target in the scene, i.e. the signal $S(k,\boldsymbol{X}_k,\boldsymbol{r})$ is described by (3.3) where $k(n) = 1$. Also we will neglect the platform instability (jitter) assuming $\boldsymbol{\delta}(n) = 0$[1]. Under these conditions

$$S(k,\boldsymbol{X}_k,\boldsymbol{r}) = A(k)h(\boldsymbol{r} - \boldsymbol{X}_k), \tag{5.2}$$

where $h(\boldsymbol{r})$ is a normalized sensor function and $A(k)$ is an unknown signal intensity.

We further define $\boldsymbol{Z}^k(\boldsymbol{r}) = (Z(1,\boldsymbol{r}),\ldots,Z(k,\boldsymbol{r}))$ to be the concatenation of all measurements up to time t_k in the space point $\boldsymbol{r}$.

The expected range of possible "behaviors" of the target is modeled as a Markov process. Often this process may be well described by a randomly perturbed multi-dimensional linear or nonlinear dynamical system

$$\dot{\boldsymbol{X}}_t = f(\boldsymbol{X}_t) + \sigma\dot{W}_t, \qquad t \geq 0,$$

where $\dot{W}_t$ is another noise process that describes uncertain and unpredictable target motion; $f(\cdot)$ is a known function. Modeling of the state and observation is normally based upon physics. The models for the state process usually represent the *a priori* knowledge about physical, tactical, etc. characteristics of the target while the observation model is based upon the physics of sensors, clutter structure, operational environment, and so forth.

[1]This is the case if the instability is compensated either by electro-mechanical stabilizators or by estimating.

5.2. **Optimal Spatio-Temporal Nonlinear Filtering: The Basic Algorithm.** The output of the ONF is the functional time series $\{\pi_k(\boldsymbol{r})\}_{k\geq 1}$ of joint posterior densities (JPD) which measure the likelihood at the time moments t_k that the vector of target features parametrized by $\boldsymbol{X}_k$ is in close proximity of the grid point $\boldsymbol{r}$.

Computation of the JPD usually splits into two separate procedures: on every time step, spatio-temporal filtering for clutter removal is done first (see Section 4), and then the spatio-temporal nonlinear filtering is performed to estimate the target location. After clutter suppression procedure the clutter is decorrelated. We thus can include it into the noise component $V(k,\boldsymbol{r})$ (see (5.1)) and concentrate only on the second procedure.

It is a standard fact (see e.g. [24], [42]) that the JPD is given by the formula

$$\pi_k(\boldsymbol{r}) = \frac{p_k(\boldsymbol{r})}{\int p_k(\boldsymbol{r})\,d\boldsymbol{r}}, \tag{5.3}$$

where the function $p_k(\boldsymbol{r})$, called the *unnormalized filtering density* (UFD), propagates in time according to the following recursive equation of the predictor-corrector type:

$$p_k(\boldsymbol{r}) = \exp\{(R^{-1}h(\boldsymbol{r},\cdot), R^{-1}Z(k,\cdot)) - \frac{1}{2}\left|R^{-1}h(\boldsymbol{r},\cdot)\right|^2\}T_k p_{k-1}(\boldsymbol{r}), \quad k=1,2,\dots, \tag{5.4}$$

where R is the covariance of the observation noise V, and the predictor $T_k p_{k-1}(\boldsymbol{r})$ is a solution of the Fokker-Planck-Kolmogorov equation corresponding to the state process subject to the initial condition $p_{k-1}(\boldsymbol{r})$, i.e. $u(t,\boldsymbol{r}) = T_t p_{k-1}(\boldsymbol{r})$ is a solution of the equation

$$\begin{aligned} &\frac{\partial u(t,\boldsymbol{r})}{\partial t} = \frac{1}{2}\sum_{i,j=1}^{r} a_{i,j}\frac{\partial^2}{\partial r_i \partial r_j}u(t,\boldsymbol{r}) - \sum_{i=1}^{m}\frac{\partial}{\partial r_i}(f_i(\boldsymbol{r})u(t,\boldsymbol{r})), \quad t\in(t_k, t_{k-1}], \\ &u(0,\boldsymbol{r}) = p_{k-1}(\boldsymbol{r}), \end{aligned} \tag{5.5}$$

where $\{a_{i,j}\}$ is the covariance of the state noise and m is the dimensionality of $\boldsymbol{X}_t$. The posterior moments of $\boldsymbol{X}_k = \boldsymbol{X}_{t_k}$ can be computed now by integration of the respective polynomials against the JPD $\pi_k(\boldsymbol{r})$. For example, the best mean square estimate of $\boldsymbol{X}_k$,

$$\widehat{\boldsymbol{X}}_k = \int \boldsymbol{r}\pi_k(\boldsymbol{r})\,d\boldsymbol{r}.$$

Recall that we consider a general problem assuming that $\boldsymbol{r}$ belongs to a $m-$dimensional Euclidean space (phase space) $\mathbb{R}^m$ (our main concern in IRST problem is $m=4$ if the problem is solved in the initial "resolution" space (angles and angular velocities) or $m=2$ if trajectories are projected on the plain).

Now, let $\{e_\ell(\boldsymbol{r})\}_{\ell\in N}$ be a complete orthonormal system in $L^2(\mathbb{R}^m)$. Projecting (5.1) on this basis, we can rewrite the observation process in the coordinate form:

$$Z_k^\ell = h^l(X_{t_k}) + V_k^\ell, \quad \ell = 1,2,\dots, \quad k=1,2,\dots,K, \tag{5.6}$$

where

$$Z_k^\ell = \int_{\mathbb{R}^m} Z_k(\boldsymbol{r})e_\ell(\boldsymbol{r})\,d\boldsymbol{r}, \quad h^\ell(\boldsymbol{y}) = \int_{\mathbb{R}^m} h(\boldsymbol{y},\boldsymbol{r})e_\ell(\boldsymbol{r})\,d\boldsymbol{r}, \quad V_k^\ell = \int_{\mathbb{R}^m} V_k(\boldsymbol{r})e_\ell(\boldsymbol{r})\,d\boldsymbol{r}$$

and V_k^ℓ are independent Gaussian random variables.

Let $F(\cdot)$ be a given function such that $\boldsymbol{E}\,|F(\boldsymbol{X}_t)|^2 < \infty$, $\forall t \geq 0$. In general, the optimal nonlinear filtering deals with obtaining the best mean square estimate of $F(X_{t_k})$, given the measurements $\boldsymbol{Z}^k(\boldsymbol{r}) = (Z_1(\boldsymbol{r}), \ldots, Z_k(\boldsymbol{r}))$, $\boldsymbol{r} \in \mathbb{R}^m$. This estimate is called the *optimal filter* and will be denoted $\hat{F}(k)$.

By T_t denote the solution operator for the Fokker-Planck-Kolmogorov equation corresponding to the process $\boldsymbol{X}$. The optimal filter $\hat{F}(k) = \boldsymbol{E}[F(\boldsymbol{X}_{t_k}) \mid \boldsymbol{Z}^k]$ is given by the formula

$$\hat{F}(k) = \frac{\int_{\mathbb{R}^m} p_k(\boldsymbol{r}) F(\boldsymbol{r})\, d\boldsymbol{r}}{\int_{\mathbb{R}^m} p_k(\boldsymbol{r})\, d\boldsymbol{r}}, \tag{5.7}$$

where the UFD $p_k(\boldsymbol{r})$ was introduced above. This density obeys the recursive relation

$$\begin{aligned} &p_0(\boldsymbol{r}) = \pi(\boldsymbol{r}), \\ &p_k(\boldsymbol{r}) = \exp\left(\Delta \sum_{\ell=1}^{\infty} h^\ell(\boldsymbol{r}) Z_k^\ell - \frac{\Delta}{2} \sum_{\ell=1}^{\infty} |h^\ell(\boldsymbol{r})|^2\right) T_\Delta p_{k-1}(\boldsymbol{r}), \quad k = 1, 2 \ldots . \end{aligned} \tag{5.8}$$

Additional notation:

$$q_{mn} = \langle e_m, T_\Delta e_n \rangle, \quad q_{mn}^\ell = \langle h^\ell e_m, T_\Delta e_n \rangle, \quad q_{mn}^{\ell j} = \langle h^\ell h^j e_m, T_\Delta e_n \rangle,$$

$$\psi_m(0) = \langle p, e_m \rangle, \quad F_m = \langle F, e_m \rangle,$$

where $\langle F, g \rangle = \int_{\mathbb{R}^m} F(\boldsymbol{r}) g(\boldsymbol{r})\, d\boldsymbol{r}$. Note that these quantities can be computed off-line before any observations become available.

Let $\overline{L} = \{\ell : \ell \leq M\}$ and $\overline{F}_k$ be the approximation to the optimal filter $\hat{F}(k)$.

We propose the following algorithm for computing the optimal nonlinear filter $\hat{F}(k)$, which will be called the Spectral Separation Scheme or the S^3 algorithm.

1. ***Set a cut-off level M for the number of basis elements.***
2. ***Before the observations become available compute:***

$$\overline{p}_0(\boldsymbol{r}) = \sum_{\ell \leq M} \psi_\ell(0) e_\ell(\boldsymbol{r}) \quad \text{and} \quad \Psi_0[f] = \sum_{\ell \leq M} \psi_\ell(0) F_\ell.$$

3. ***When the measurements (Z_k^ℓ), $k = 1, 2, \ldots, K$, $\ell \in L \subset \overline{L}$ become available:***

a): *compute recursively*

$$\psi_m^L(0) = \psi_m(0), \quad \psi_m^L(k) = \sum_{n \leq M} Q_{mn}(Z_k^L) \psi_n^L(k-1), \quad k = 1, 2, \ldots;$$

$$Q_{mn}(Z_k^L) = q_{mn} + \Delta \sum_{\ell \in L} Z_k^\ell q_{mn}^\ell + \Delta^2 \sum_{\ell \neq j; \ell, j \in L} Z_k^\ell Z_k^j q_{mn}^{\ell j} + \frac{\Delta}{2} \sum_{\ell \in L} \big((Z_k^\ell)^2 \Delta - 1\big) q_{mn}^{\ell\ell};$$

b): *compute*

$$\overline{p}_k(\boldsymbol{r}) = \sum_{\ell \leq M} \psi_\ell^L(k) e_\ell(\boldsymbol{r}), \quad \Psi_k[F] = \sum_{\ell \leq M} \psi_\ell^L(k) F_\ell, \quad \overline{F}_k = \Psi_k[F]/\Psi_k[1].$$

The block diagram of the Spectral Separation Scheme (S^3 algorithm) is shown in Figure 2. It is clear that the above algorithm allows for an adaptive, dynamic interplay between data and computations. Indeed, it does not involve solving of PDE's on-line and is recursive in time. Furthermore, it is also *spatially recursive.* These properties are very important since they allow *adaptive sequential multi-resolution filtering* to be performed. Observe that in order to achieve the effect of multi-resolution one has to use local hierarchical bases $\{e_l\}$, for example *wavelet bases* [25].

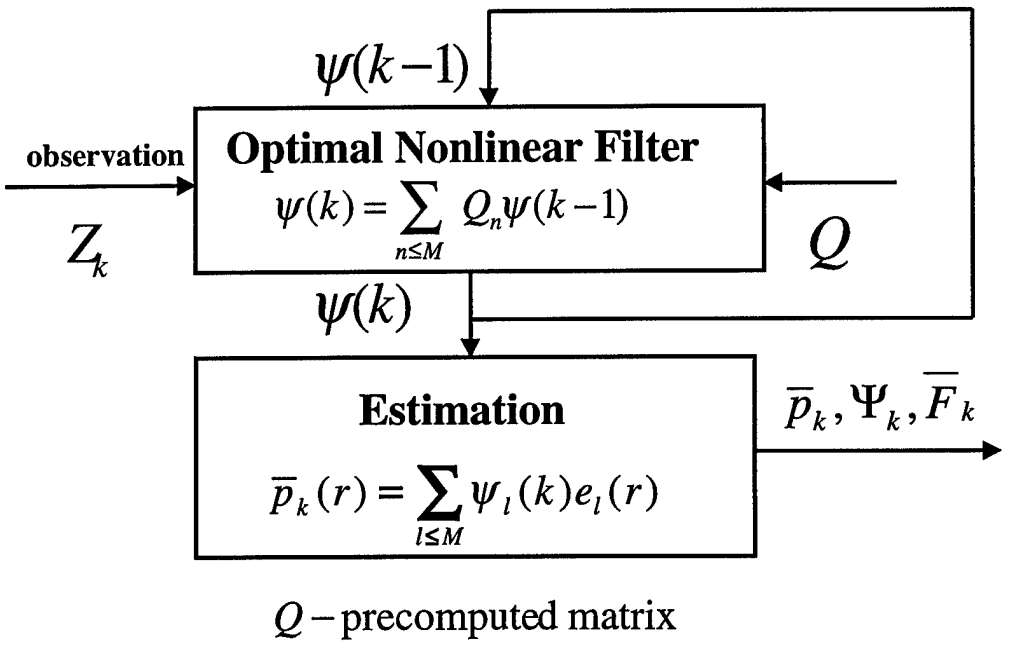


Figure 2. S^3 algorithm (on-line part)

5.3. **Multi-Dimensional Spatio-Temporal Matched Filter as a Special Case of Nonlinear Filter.** For the sake of simplicity let us consider the basic formula for updating the filtering density in the optimal nonlinear filter in the 3−D case where the target motion is projected on the plain. The modification for the $(m+1)$−D case with arbitrary m, particularly for $m=4$, is straightforward. The grid space will be denoted by $\{x_{ij}\}$. Suppose that the solution of the Fokker-Plank-Kolmogorov equation (the mean motion dynamics of the target) is given by the operator T_Δ and observations are made at time moments $t_k = k\Delta$ on the observation space grid $\{x_{ij}\}$ with additive Gaussian white noise:

$$Z_{ij}(t_k) = h(X_{t_k} - x_{ij}) + \sigma\xi_{ij}(t_k), \tag{5.9}$$

where $h(\cdot)$ is the target signature signal and $\xi_{ij}(t_k)$ are i.i.d. standard Gaussian variables.

Then by (5.4) the updating formula is

$$p_k(x) = \exp\Big\{\frac{1}{\sigma^2}\sum_{i,j} h(x - x_{ij})Z_{ij}(t_k) - \frac{1}{2\sigma^2}\sum_{i,j} h(x - x_{ij})^2\Big\}T_\Delta p_{k-1}(x)\,.$$

In the above formula, the part $T_\Delta p_{k-1}(x)$ is the *prediction part*, i.e. the best possible guess (based on known target trajectory dynamics), what the probabilities to find the target in various areas of the observation region would be if there were no observations available at the time moment t_k. The exponent part is the *correction part*, i.e. after receiving the observation at time t_k, the prediction part is amplified or attenuated, depending on the expression inside the exponent.

One notable fact is that the amplification or attenuation is directly related to the output of the well-known in engineering community *single frame matched filter* (spatial (2−D) matched filter)

$$m_k(x) = \sum_{i,j} h(x - x_{ij}) Z_{ij}(t_k).$$

Indeed, the filtering densities are amplified only at those points where the output of the matched filter exceeds the value of

$$\tfrac{1}{2}\mathrm{E}^2 = \frac{1}{2}\sum_{i,j} h(x - x_{ij})^2,$$

and attenuated elsewhere. Here E^2 is the energy of the signal which of course does not depend on x.

This establishes a strong connection between a spatial matched filter and the optimal nonlinear filter. Even more importantly, the optimal nonlinear filter is equivalent to the *spatio-temporal matched filter* (or *assumed velocity matched filter*) if the movement of the target occurs along a deterministic trajectory. In that case the operator T_Δ degenerates into an operator that shifts the values of the function along the known deterministic trajectory. For simplicity, let us assume that the target moves in a straight line with known constant velocity v. Then

$$T_\Delta p(x) = p(x - v\Delta),$$

and it is easy to see that for any $k \geq 1$

$$p_k(x) = c_k \exp\Big\{ \frac{1}{\sigma^2} \sum_{\ell=0}^{k-1} \sum_{i,j} h(x - x_{ij} - \ell v\Delta) Z_{ij}(t_k) - \frac{1}{2\sigma^2} \sum_{\ell=0}^{k-1} \sum_{i,j} h(x - x_{ij} - \ell v\Delta)^2 \Big\} p_0(x - kv\Delta).$$

Notice that the first part in the exponent in the above formula,

$$M_k(x) = \sum_{\ell=0}^{k-1} \sum_{i,j} h(x - x_{ij} - \ell v\Delta) Z_{ij}(t_k)\,,$$

exactly coincides with the usual assumed velocity matched filter and the second part

$$\frac{1}{2\sigma^2} \sum_{\ell=0}^{k-1} \sum_{i,j} h(x - x_{ij} - \ell v\Delta)^2 = \frac{\mathrm{E}^2}{2\sigma^2} k$$

is nothing but the accumulated SNR. Thus, if we assume that the initial density is a delta-function, then statistical inference base on $M_k(x)$ and $p_k(x)$ will be the same. Particularly, the points of maxima of $M_k(x)$ and $p_k(x)$ coincide.

A similar argument can be applied to any dimensions. (Recall that in general we are interested in the $(m+1)$−D case, $\boldsymbol{r} = (r_1, \ldots, r_m)$, $(m+1)$−th component is time.) The final result is

exactly the same: *in case of deterministic trajectory the* $(m+1)-D$ *matched filter and the optimal spatio-temporal nonlinear filter coincide with the accuracy to the deterministic transformation.*

We stress that the developed optimal spatio-temporal nonlinear filter does more than the spatio-temporal matched filter does. It is a generalization of the multi-dimensional spatio-temporal matched filter and coincides with it for targets that move along deterministic trajectories. In general, however, the ONF predicts the trajectory and allows us to align frames coherently even for acutely maneuvering targets when the $(m+1)-D$ ***matched filter fails.***

6. Track Appearance/Disappearance Detection

6.1. **Preliminaries.** A problem of detecting target tracks that occur (appear and disappear) at *a priori* unknown points in time is a typical abrupt change detection problem. The change-point detection problem has been actively researched during the last three decades (see, e.g., [22, 34, 37, 41, 46, 44, 52, 53, 54] and references therein). Heuristic procedures such as Shewhart's charts and some modifications appeared in the late twenties and early thirties. Detection procedures, which are in current use, were initiated by Girshik and Rubin [22] and Page [34] for the problem of detecting a change in a mean of i.i.d. Gaussian sequence and by Shiryaev [44] in the general (but i.i.d.) case. There are two major competitive procedures: the Shiryaev-Roberts-Girshik-Rubin algorithm [41, 37, 54, 44] and the Page's (or CUSUM – cumulative sum) procedure [34, 46, 55][2].

There are two major disadvantages to both algorithms:

1. Only one change in data is assumed, i.e. in our context the moment of target disappearance is ignored.
2. The aforementioned change-point detection procedures have optimal properties only in the i.i.d. case in the following specific sense: they minimize the mean detection delay under constraints on the mean time between false alarms for exactly specified pre-change and post-change distributions.

These drawbacks make the conventional change-point detection procedures impractical in multi-target surveillance problems that involve non-i.i.d. observations and when one needs not only to detect a particular target (with unknown position) but also to discriminate between false alarms and true tracks. One appropriate criterion for our purposes is to fix probabilities of false alarm and detection in a fixed size sliding window and to minimize the detection delay. Or, alternatively, one may require to detect the track with fixed or maximum probability during the fixed time and with minimum detection delay. Also it is imperative to design the decision statistics that will account for unknown target location and moment of its disappearance. Otherwise the problem is meaningless. Neither of the above mentioned algorithms may be directly applied in this context. However, we will show that the CUSUM-type procedure and the quasi-Bayesian (Shiryaev-Roberts-Girshik-Rubin type) procedure with specially designed thresholds are useful

[2]See also [40, 54] for reasonable modifications and ideas that are useful for surveillance systems.

tools, especially in the problem of detection of a single target that appears and disappears at unknown times. Moreover, it turns out that the thresholds can be chosen such that the predefined false alarm rate is fixed for a large class of noise and signal models, not necessarily for the i.i.d. models. This innovation is very important for our applications, since it removes the restrictive i.i.d. assumption, which is typical for previous work in change-point detection. In particular, it allows us to design thresholds for adaptive algorithms.

Now we summarize the basic requirements that should be satisfied in realistic problems of detection of target's track appearance/disappearance:

- detection algorithms should be invariant relative to the completely unknown moments of abrupt change (the prior distributions of appearance/disappearance moments are unknown);
- detection algorithms should be adaptive and use estimates of unknown target location based on TbD;
- the frequency of false detections or the false alarm probability is limited at the specified level;
- the probability of correct detection should be maximal during the fixed time interval;
- the mean detection delay should be made as small as possible.

We propose the sequential algorithms that meet all these requirements (and even more, see Algorithm 3 in Section 6.5). The algorithms are based either on the generalized CUSUM statistic or on the quasi-Bayesian statistic. These statistics are, in essence, score functions that characterize the likelihood ratio of hypotheses on signal presence and absence. Specifically, the generalized CUSUM statistic is the likelihood ratio maximized over unknown moments of appearance and disappearance and the latter statistic is the average likelihood ratio with respect to flat (improper) prior distributions of these moments. The statistics are computed either in a sliding window of a fixed size or on the semiinfinite interval (up to the current moment). In both cases the decision statistics are compared with thresholds that depend on a given false alarm rate (probability of false alarm or frequency of false alarms). It turns out that in many cases the generalized CUSUM procedure may be reduced to the conventional CUSUM without loss of statistical performance. To be precise, if the major goal is to detect the fact of target's appearance (but not the fact of its disappearance), than the unknown moment of target's disappearance should be considered as a nuisance parameter. In this case the CUSUM is the optimal algorithm, i.e. the unknown moment of target's disappearance does not affect the structure of the algorithm, but only its performance (see Section 6.3 and Section 6.4 for details). The same is true for the quasi-Bayes procedure. If, however, one has to detect multiple targets, then it is important to detect both track appearance and track disappearance (as soon as possible). Moreover, in this case it is also important to have a special logic to discriminate between false alarms and true tracks. In this "multi-target context" the conventional CUSUM-type procedures do not solve the problem but still may be used as a part of more sophisticated algorithms (see Section 6.5 for some discussion).

It should be mentioned that the problem of signal detection with random appearance and disappearance times in ideal conditions (complete prior information) was solved by Tartakovsky

[56]. The results of this work show that the optimal Bayes solution requires, in general, computation of the 5−dimensional sufficient statistic and is completely impractical. Thus even the availability of the required prior information, particularly knowledge of prior distributions of appearance/disappearance moments (which are usually unknown), does not help in practice.

6.2. **Problem Formulation and Model Description.** Let Y_k, $k = 1, \ldots, n$, denote the data obtained up to the current time n. Generally, the data $Y_k = \{\boldsymbol{Z}_k(\boldsymbol{r}_i), i = 1, \ldots, N\}$ represent the whole "spatial set" collected from the frame of observations (after preprocessing) at time k or the part of these data (in some "channels"). If the clutter removal procedure is first performed, then (after spatio-temporal filtering) the clutter is decorrelated. Thus, it is reasonable to assume that $\{Y_k\}$ are i.i.d. with the density $p_1(y)$ if the target is present and with the density $p_0(y)$ if it is absent[3]. The density $p_1(Y_k \mid \theta_k)$ depends on (in general unknown) vector parameter θ_k, which characterizes the spatial location of the target at moment k. First, for the sake of simplicity we assume that the target dynamics is known and hence the θ_k is fixed and known. The exact models with the account of TbD and corresponding adaptive algorithms will be defined later on in Section 6.8 and Section 6.9 (see also Section 6.6 in the general case).

By $\Lambda_k = \log \frac{p_1(Y_k)}{p_0(Y_k)}$ denote the log-likelihood ratio (LLR) for the single observation Y_k and by λ and γ the unknown moments of target appearance and disappearance, respectively. Under our assumptions the data $Y_1, \ldots, Y_{\lambda-1}$ are i.i.d. according to the density $p_0(y)$, $Y_\lambda, \ldots, Y_\gamma$ are i.i.d. according to $p_1(y)$ and $Y_{\gamma+1}, Y_{\gamma+2}, \ldots, n$ again follow $p_0(y)$. We assume nothing about λ and γ except for the natural constraint $\gamma > \lambda$. In the language of hypotheses testing the problem of the track detection may be formulated as testing of the hypotheses

"$H_{1,l,m}$: track appears at $\lambda = l$ and disappears at $\gamma = m$, $\quad 1 \le l \le n,\ m \ge l + 1$";

"H_0 : track does not appear, i.e. $\lambda \notin [1, n]$".

Alternatively, the hypotheses may be written in the form

$$\text{"}H_{1,\lambda,\gamma} : p(Y_1^n) = p_{\lambda,\gamma}(Y_1^n) = \prod_{k=1}^{\lambda-1} p_0(Y_k) \times \prod_{k=\lambda}^{\gamma} p_1(Y_k) \times \prod_{k=\gamma+1}^{n} p_0(Y_k)\text{"}\,,$$

$$\text{"}H_0 : p(Y_1^n) = p_0(Y_1^n) = \prod_{k=1}^{n} p_0(Y_k)\text{"}.$$

6.2.1. ***Generalized Likelihood Ratio Statistic.*** By L_n denote the generalized LLR,

$$L_n := \max_{\lambda,\gamma} \log \frac{p_{\lambda,\gamma}(Y_1^n)}{p_0(Y_1^n)} = \max_{\substack{1 \le \lambda \le n \\ \gamma \ge \lambda+1}} \sum_{k=\lambda}^{\gamma} \Lambda_k.$$

[3]The derived algorithms can be easily modified and generalized for more general models that include correlated and non-homogeneous observations, see comments below and Section 6.7.

It is easily shown that the statistic L_k satisfies the following system of recursive relations:

$$L_k = \max\big(L_{k-1}, U_k\big) = L_{k-1} + \big(U_k - L_{k-1}\big)^+, \tag{6.1}$$

$$U_k = \Lambda_k + U_{k-1}^+, \qquad k = 1, \dots, n, \tag{6.2}$$

with the initial conditions $L_0 = U_0 = 0$. Here y^+ denotes a non-negative part of y, i.e. $y^+ = \max(0, y)$. Indeed,

$$L_n = \max_{\gamma \geq 1} \left(\max_{1 \leq \lambda \leq \gamma} \sum_{k=\lambda}^{\gamma} \Lambda_k \right) = \max(U_1, \dots, U_n) = \max(L_{n-1}, U_n)\,,$$

where for any $k \geq 1$

$$U_k =: \max_{1 \leq \lambda \leq k} \sum_{s=\lambda}^{k} \Lambda_s = \max \left[\Lambda_k, \max_{1 \leq \lambda \leq k-1} \left(\Lambda_k + \sum_{s=\lambda}^{k-1} \Lambda_s \right) \right] = \Lambda_k + \max(0, U_{k-1})\,.$$

6.2.2. ***Generalized Average Likelihood Ratio Statistic.*** Introduce the following statistic

$$G_n := \max_{\gamma} \int_0^{\gamma} \frac{p_{\lambda,\gamma}\big(Y_1^n\big)}{p_0\big(Y_1^n\big)}\, d\lambda = \max_{\gamma \geq 1} \sum_{\lambda=1}^{\gamma} \prod_{k=\lambda}^{\gamma} \exp(\Lambda_k)\,.$$

It is easy to see that

$$G_n = \max(R_1, R_2, \dots, R_n) = \max(G_{n-1}, R_n), \quad G_0 = 0,$$

where $R_n = \sum_{\lambda=1}^{n} \prod_{k=\lambda}^{n} e^{\Lambda_k}$. The statistic R_n may be interpreted as the likelihood ratio averaged over the flat (uniform) distribution of the moment of track appearance.

It is easy to show that $\{R_n\}$ obeys the recursive relation

$$R_n = \exp(\Lambda_n)(1 + R_{n-1}), \quad R_0 = 0\,. \tag{6.3}$$

Both statistics, L_n and G_n, can be used for testing the above hypotheses either in the fixed interval or sequentially.

6.3. **Detection Algorithm 1: Fixed Sliding Window.** The hypotheses $H_{1,\lambda,\gamma}$ and H_0 are tested in a sliding window of length T at each current time instant n, i.e.

$$\text{“}H_{1,\lambda,\gamma} : p(Y_{n-T+1}^n) = p_{\lambda,\gamma}(Y_{n-T+1}^n) = \prod_{k=n-T+1}^{\lambda-1} p_0(Y_k) \times \prod_{k=\lambda}^{\gamma} p_1(Y_k) \times \prod_{k=\gamma+1}^{n} p_0(Y_k)\text{”}\,,$$

$$\text{“}H_0 : p(Y_{n-T+1}^n) = p_0(Y_{n-T+1}^n) = \prod_{k=n-T+1}^{n} p_0(Y_k)\text{”}.$$

In this case the statistics $L_{T,n}$ and $G_{T,n}$ are determined by

$$L_{T,k} = \max\big(L_{T,k-1}, U_k\big), \tag{6.4}$$

$$U_k = \Lambda_k + U_{k-1}^+, \qquad k = n - T + 1, \dots, n, \tag{6.5}$$

and

$$G_{T,k} = \max\big(G_{T,k-1}, R_k\big), \tag{6.6}$$

$$R_k = e^{\Lambda_k}(1 + R_{k-1}), \qquad k = n - T + 1, \ldots, n, \tag{6.7}$$

with the initial conditions $L_{T,n-T} = U_{n-T} = G_{T,n-T} = R_{n-T} = 0$.

According to the adaptive Bayesian approach, the procedures of testing the hypothesis H_1 against H_0 have the form

$$d^L_{T,n} = \begin{cases} 1 & \text{if } L_{T,n} \geq a, \\ 0 & \text{otherwise,} \end{cases} \tag{6.8}$$

$$d^G_{T,n} = \begin{cases} 1 & \text{if } G_{T,n} \geq B, \\ 0 & \text{otherwise,} \end{cases} \tag{6.9}$$

where a, B are thresholds that are defined on the basis of the given probability of false alarm and $d = 1$ stands for the decision on target presence while $d = 0$ is the decision on its absence. It should be emphasized that in this subsection we pursue the goal of detection of the target track appearance regardless of the possibility of its disappearance in the analyzed observation interval. In other words, the unknown moment of target disappearance is considered as a nuisance parameter. A different setting when both target's appearance and disappearance should be detected will be considered in the next two sections.

We now notice the following remarkable property of the tests (6.8) and (6.9). Since

$$L_{T,n} = \max_{n-T+1 \leq k \leq n} U_k \,,$$

we have

$$\{d^L_{T,n} = 0\} = \{U_k < a \text{ for all } \; n - T + 1 \leq k \leq n\} \,.$$

This shows that the non-sequential test (6.8) is equivalent to the sequential test which is defined by the stopping time

$$\tau_a(T,n) = \min\{n - T + 1 \leq k \leq n : U_k \geq a\}, \quad \tau_a(T,n) = \infty \text{ if no such } k. \tag{6.10}$$

If $\tau_a(T,n) < \infty$, then the hypothesis H_1 is accepted (target track is present), while if $\tau_a(T,n) = \infty$, the hypothesis H_0 (target is absent) is accepted. The same argument is applied to show that the non-sequential test (6.9) is equivalent to the sequential one

$$\tilde{\tau}_B(T,n) = \min\{n - T + 1 \leq k \leq n : R_k \geq B\}, \quad \tilde{\tau}_B(T,n) = \infty \text{ if no such } k. \tag{6.11}$$

The sequential procedures (6.10) and (6.11) allow us to achieve exactly the same probabilities of false alarm and target missing as the initial tests (6.8) and (6.9), respectively. In addition, they have an advantage: the time delay in signal detection is less. See Figure 3 for graphical explanation of the above argument.

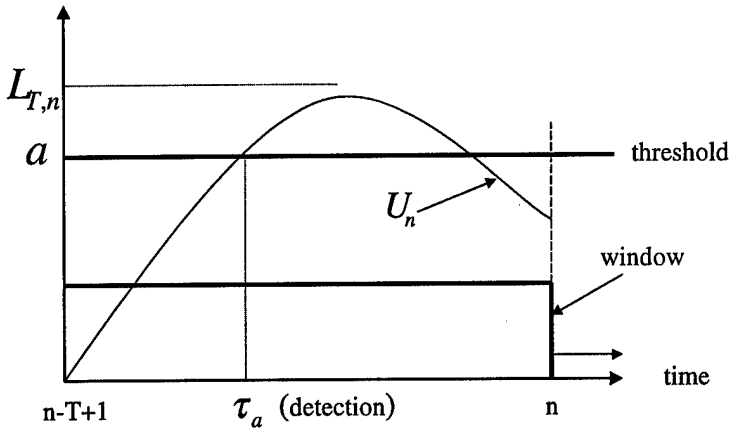


Figure 3. Detection in the sliding window: the statistic U_n exceeds the threshold at random moment τ_a, which is less than the size of the sliding window T.

We also observe the following interesting fact. The statistic U_k is nothing but the CUSUM statistic, which is defined in the standard CUSUM procedure as

$$U_k = \max_{\lambda \le k} \sum_{s=\lambda}^{k} \Lambda_s,$$

i.e. as the maximum of the LLR over unknown moment of signal appearance only. In turn, the statistic R_n is nothing but the quasi-Bayesian statistic, which is the limit form of the Bayesian statistic (the Shiryaev-Roberts-Girshik-Rubin statistic). This shows that the presence of an additional nuisance parameter – moment of signal disappearance γ, does not affect the final structure of the test procedures. But it certainly does affect the detection performance, since the accumulated SNR becomes less as γ decreases.

Another possible modification of the algorithm is to use the statistic $\tilde{U}_k$ with reflection from zero barrier,

$$\tilde{U}_k = (\Lambda_k + U_{k-1})^+ , \quad k \ge 1, \qquad \tilde{U}_0 = 0 , \tag{6.12}$$

in place of U_k in (6.10). Obviously, these two tests have the same performance as long as the threshold is positive. Indeed, it is easy to see that the trajectories of the statistics U_k and $\tilde{U}_k$ coincide in the non-negative half-plain.

6.4. **Detection Algorithm 2: Fully Sequential Procedure.** We now change the problem set-up and consider the fully sequential procedure. To be specific, we assume that $k = n, n-1, \ldots$ and the decision about target presence or absence is made at each moment. This case is probably the most relevant to the IRST surveillance systems, which work for a long time periodically raising false alarms.

It is easy to see that the generalized LLR $L_{-\infty,n} := L_n$ is

$$L_n = \max(U_n, U_{n-1}, \cdots),$$

where U_k is defined in (6.5). Formally the sequential algorithm that solves the problem is identified with the stopping time

$$\tau_a = \inf\{n : L_n \geq a\}.$$

The stopping time τ_a can be rewritten as

$$\tau_a = \inf\{n : U_n \geq a\}, \tag{6.13}$$

which is exactly the CUSUM procedure. Indeed, the decision on target absence ($d_n = 0$), which is equivalent to observation continuation at time n, is made when

$$\{L_n < a, L_{n-1} < a, \cdots\} = \{U_n < a, U_{n-1} < a, \cdots\}.$$

Thus, again prior uncertainty with regard to the moment of disappearance does not affect the structure of the detection algorithm (if γ is considered as a nuisance parameter) but it does affect its performance.

It is important to understand that the algorithm (6.13) requires additional strategy after the decision $d = 1$ (target appeared) is made. Since we expect multiple signal appearances, one may not simply stop observations after this decision is made. One way to handle this problem is to set $U_n = 0$ when $\tau_a = n$ and to start all over again (immediate renewal). Then the algorithm is immediately ready to detect the next target. In this case the CUSUM statistic is modified (compared to (6.5)) as follows

$$U_n = \Lambda_n + U_{n-1}\mathbb{1}_{\{0<U_{n-1}<a\}}, \tag{6.14}$$

where $\mathbb{1}_{\{Y\}}$ is the indicator of the set Y. This modification has, however, one drawback: it may (and usually does) lead to multiple detections of the same signal and hence requires additional identification logic.

Another reasonable way is to use the detection rule (6.13) with pure CUSUM (6.5) supplemented by the following logic for new target detection. Let ν_a be the first time such that U_n goes below the threshold a after exceeding. To be precise,

$$\nu_a = \inf\{n > \tau_a : U_n < a\}.$$

Then for $\tau_a \leq k \leq \nu_a - 1$, one confirms the presence of the same target, while if for some $m \geq 1$, $U_{\nu_a+m} \geq a$, then we make the decision that a new target track appears at time $\nu_a + m$. In other words, if the $i-$th target was detected at time $\tau_a^{(i)}$ and confirmed till $\nu_a^{(i)} - 1$, then the $(i+1)-$th target is detected at

$$\tau_a^{(i+1)} = \inf\{n > \nu_a^{(i)} : U_n \geq a\}$$

and its track is confirmed till the moment $\nu_a^{(i+1)} - 1$, where

$$\nu_a^{(i+1)} = \inf\{n > \tau_a^{(i+1)} : U_n < a\}.$$

We also note that one may replace the statistic (6.14) by the statistic

$$\tilde{U}_n = (\Lambda_n + U_{n-1})\mathbb{1}_{\{0<\Lambda_n+U_{n-1}<a\}}. \tag{6.15}$$

This replacement does not affect performance.

Applying a similar argument to the quasi-Bayesian statistic G_n, we obtain that the stopping rule

$$\tau_B = \inf\{n : G_n \geq B\}$$

can be rewritten as

$$\tau_B = \inf\{n : R_n \geq B\}. \tag{6.16}$$

In the case of immediate renewal the average likelihood ratio R_n is computed recursively according to the formula

$$R_n = \exp(\Lambda_n)\left(1 + R_{n-1}\mathbb{1}_{\{R_{n-1}<B\}}\right). \tag{6.17}$$

The same logic as above can be used to detect not only the fact of target appearance but also its disappearance. However, in the case of multiple targets it is better to consider a more complex set of hypotheses that would include at least three alternatives: the last signal did not appear, appeared but is still present, appeared and disappeared. This case is considered in the next section.

We stress once again that the statistic R_n may be interpreted as the average likelihood ratio over the uniform distribution of the λ on the interval $[1, n]$, while the $\exp(U_n)$ is the maximum of the same likelihood ratio over $\lambda \in [1, n]$.

The method (6.16) has an advantage over the generalized CUSUM algorithm – the statistic $\tilde{R}_n = R_n - n$ is a zero-mean martingale with respect to $\boldsymbol{P}_0$ regardless of the i.i.d. assumption on the observations. As a result, if one sets $B = 1/\mathrm{Fr}$, then the frequency of false alarms is upper bounded by the prespecified value Fr (see Section 6.7.1 for details). In other words, the threshold B is easily estimated for a large class of models. There is no similar result for the CUSUM statistic U_n.

6.5. **Algorithm 3: Joint Detection of Track Appearance and Disappearance.** Consider an extended decision making process assuming that after each decision on target disappearance the previous data are discarded. Specifically, we accept the following logic:

1. If the decision that the $i-$th target disappeared is made ($d_i = \mathrm{DA}$), then one of the two decisions may be made – the $(i+1)-$th (new) "target did not appear" ($d_{i+1} = \mathrm{NA} = 0$) or the $(i+1)-$th "target appeared and is still present" ($d_{i+1} = \mathrm{A\&P}$).
2. If the decision $d_{i+1} = 1$ that the $(i+1)-$th target appeared is made, then one of the two decisions may be made – "target is still present" ($d_{i+1} = \mathrm{P}$) or "target disappeared" ($d_{i+1} = \mathrm{DA}$).

The statistics L_n and G_n are then the likelihood (relative to H_0 – target is absent at all) of the composite (and combined) hypothesis H that includes two sub-alternatives:

"H : target appeared and is still present ($H_{\mathrm{A\&P}}$) + target appeared and disappeared ($H_{\mathrm{A\&D}}$)".

Clearly, at time moment n the likelihood of the subalternative $H_{\text{A\&P}}$ is determined by the statistic U_n, while the likelihood of the subalternative $H_{\text{A\&D}}$ by $L_{n-1} = \max_{k \le n-1} U_k$. Thus the structure of the decision making algorithm may be as follows.

After the decision $d_i = \text{DA}$ (the $i-$th target disappeared) is made, the statistic U_n is formed with the null initial condition. This statistic is compared (at each step) with the threshold a. If $U_n < a$, the observation is continued. Otherwise ($U_n \ge a$), the decision $d_{i+1} = 1$ (the $(i+1)-$th target appeared) is made, the statistic U_n is computed further and also the algorithm starts to compute the statistic L_n with the initial condition $L_{\tau_a^{(i+1)}} = U_{\tau_a^{(i+1)}}$, where $\tau_a^{(i+1)}$ is the moment of detection of the $(i+1)-$th target. The difference $\Delta_n = L_{n-1} - U_n$ is compared with another threshold b at each step $n = \tau_a^{(i+1)} + 1, \tau_a^{(i+1)} + 2, \ldots$. If $\Delta_n < b$, then the decision $d_{i+1} = \text{P}$ (still present) is made and the next time step is analyzed. If for some $n = \nu_b^{(i+1)}$, $\Delta_n \ge b$, then the decision $d_{i+1} = \text{DA}$ (target disappeared) is made, the statistic L_n is not computed, but the statistic U_n is computed for $n = \nu_b^{(i+1)} + 1, \nu_b^{(i+1)} + 2, \ldots$ with the initial condition $U_{\nu_b^{(i+1)}} = 0$. Then the whole cycle is repeated.

Thus, the track appearance of the $(i+1)-$th target is detected at the moment

$$\tau_a^{(i+1)} = \inf\{n > \nu_b^{(i)} : U_n \ge a\}$$

and its disappearance is detected at

$$\nu_a^{(i+1)} = \inf\{n > \tau_a^{(i)} : \Delta_n \ge b\}\,.$$

The value of the threshold b may be computed based on the trade-off between the error probabilities due to too early decision about target disappearance and the delay of this decision when the target really disappears.

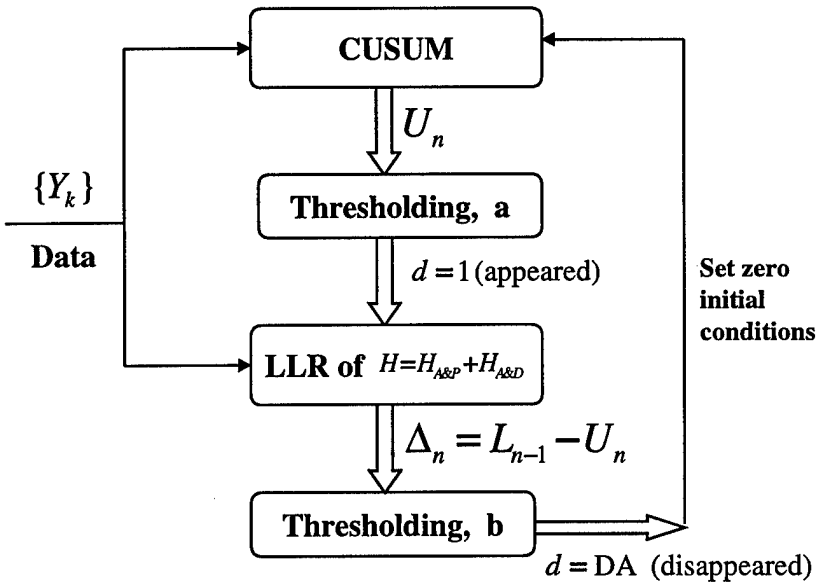


Figure 4. Block diagram of Algorithm 3

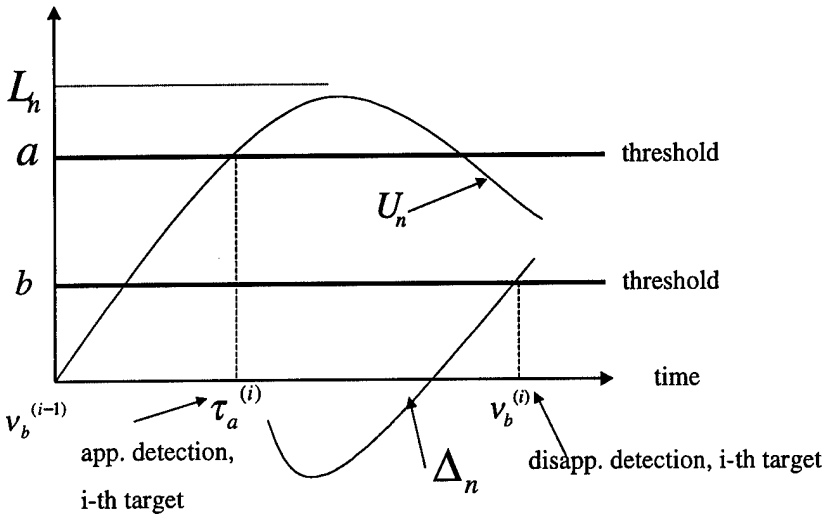


Figure 5. Detection of track appearance/disappearance by Algorithm 3

The similar decision making algorithm may be built upon the statistics R_n and G_n. Then the appearance of the $(i+1)-$th target is detected at the moment

$$\tilde{\tau}_B^{(i+1)} = \inf\{n > \tilde{\nu}_c^{(i)} : \log R_n \geq \log B\}$$

while its disappearance is detected at time

$$\tilde{\nu}_c^{(i+1)} = \inf\{n > \tilde{\tau}_B^{(i)} : \tilde{\Delta}_n \geq c\}\,,$$

where $\tilde{\Delta}_n = \log G_{n-1} - \log R_n$ with the initial condition $\tilde{\Delta}_{\tilde{\tau}_B^{(i)}} = 0$ and where c is a threshold.

The block diagram of the algorithm and the typical behavior of the decision statistics are shown in Figure 4 and Figure 5, respectively.

6.6. **Adaptive Detection Algorithms.** Recall that in general the position of the target is unknown and the density $p_1(Y_k \mid \theta_k)$ depends on the parameter θ_k that characterizes the position. Then instead of unknown θ_k one may use the estimate of the position. This estimate can be obtained by applying the TbD procedure (based on ONF). Thus the development of adaptive versions of the above detection algorithms is needed.

The statistics $U_n = U_n(\theta_1, \ldots, \theta_n)$ and $R_n = R_n(\theta_1, \ldots, \theta_n)$ are the functions of the sequence of theta's till time n. Thus the developed procedures can not be applied directly. If the θ_n can be estimated (this is the case in our system), then the natural solution is to use the statistics $\hat{U}_n = U_n(\hat{\theta}_1, \ldots, \hat{\theta}_n)$ and $\hat{R}_n = R_n(\hat{\theta}_1, \ldots, \hat{\theta}_n)$, which are computed based on the recursive formulas

$$\hat{U}_n = \Lambda_n(\hat{\theta}_n) + \hat{U}_{n-1}^{+}, \quad \hat{R}_n = e^{\Lambda_n(\hat{\theta}_n)}(1 + \hat{R}_{n-1})\,.$$

Here $\hat{\theta}_k = \hat{\theta}_k(Y_1^k)$ is an estimate of θ_k based on the previous k observations.

However, it is very difficult to evaluate the performance of the corresponding adaptive algorithms. The reason is that $\hat{U}_n$ is not a CUSUM and $\hat{R}_n$ is not an average likelihood ratio anymore. In

fact, the $\exp\{\Lambda_n(\hat{\theta}_n)\}$ is not the partial likelihood ratio, since $p_1(Y_k \mid \hat{\theta}_k(Y_1^k))$ is not a probability density. Particularly, the most important question of how to choose the thresholds remains open.

To avoid this difficulty we propose the following trick. At stage k instead of using the estimate $\hat{\theta}_k$ we propose to use the one stage delayed estimate $\hat{\theta}_{k-1}(Y_1^{k-1})$. In other words, instead of $\hat{U}_n$ and $\hat{R}_n$ we will use the statistics U_n^* and R_n^*, which satisfy the recursions

$$U_n^* = \Lambda_n^* + [U_{n-1}^*]^+, \quad n = 2, 3, \ldots, \quad U_1^* = 0; \tag{6.18}$$

$$R_n^* = e^{\Lambda_n^*}(1 + R_{n-1}^*), \quad n = 2, 3, \ldots, \quad R_1^* = 0, \tag{6.19}$$

with $\Lambda_n^* = \Lambda_n(\hat{\theta}_{n-1})$. The corresponding adaptive target detection procedures have the form

$$\tau_a^* = \inf\{n \geq 2 : U_n^* \geq a\}, \tag{6.20}$$

$$\tilde{\tau}_B^* = \inf\{n \geq 2 : R_n^* \geq B\}. \tag{6.21}$$

Since $p_1(Y_n \mid \hat{\theta}_{n-1})$ is the probability density for any Y_1^{n-1}–measurable estimate of θ_n, Λ_n^* is the log-likelihood ratio. As a result, the statistics (6.18), (6.19) preserve most of the nice properties of the former statistics U_n and R_n. In particular, $R_n^* - n$ is a $\boldsymbol{P}_0$–martingale with mean zero. This fact allows us to upper bound the frequency of false alarms in the adaptive algorithms regardless of specific pre-change and post-change distributions (see Section 6.7.1).

The block-diagram of a typical adaptive track appearance/disappearance detection algorithm that uses the estimates of target location from the ONF-based TbD scheme is shown in Figure 6.

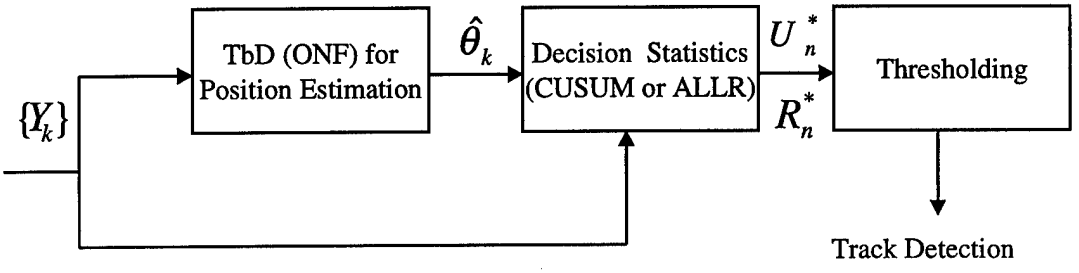


Figure 6. Block diagram of a typical adaptive detection algorithm

6.7. **Choice of Thresholds and Performance Evaluation.** In this section we give an argument and some ideas that can be used for performance evaluation of the proposed detection algorithms. We focus on the first two algorithms. The analysis of the third algorithm is more difficult and we leave this problem for the future.

In both algorithms (Algorithm 1 and Algorithm 2) it is desirable to fix the probability of false alarm P_{fa} in a given fixed interval of length T. We note that while in Algorithm 1 this is the length of the sliding window, in Algorithm 2 this is the "artificial", auxiliary parameter that is defined based on tactical conditions[4]. For the fully sequential algorithms (Algorithm 2) it is also

[4]If the dense flow of targets from a particular direction is expected, T should be chosen much less than the typical average duration of a signal. Otherwise T should be comparable with the average time of target presence.

reasonable to fix the frequency of false alarms, $\mathrm{Fr} = 1/\boldsymbol{E}_0\tau$, at the specified level $\overline{\mathrm{Fr}}$. In the next subsection we show that it may be done in the general, not necessarily i.i.d. case.

6.7.1. ***Upper Bounds for the Frequency of False Alarms.*** Consider the general case where observations can be dependent and/or non-stationary. To be specific, assume that under H_0, $p(Y_1^n) = p_0(Y_1^n)$, while if the target appears at the moment λ and is still present at time n the model is

$$p_\lambda(Y_1^n) = p_0(Y_1^{\lambda-1}) \times p_1(Y_\lambda^n \mid Y_1^{\lambda-1}),$$

where $p_0(\cdot)$ is the joint density of the noise and $p_1(\cdot)$ describes probabilistic properties of the mixture of signal and noise. By $g_\lambda^n(Y_1^n) = \frac{p_\lambda(Y_1^n)}{p_0(Y_1^n)}$ denote the likelihood ratio of the hypotheses H_λ and H_0. Obviously,

$$g_\lambda^n = g_\lambda^{n-1} \cdot g_n^n \qquad \text{where} \quad g_\lambda^k = \frac{p_1(Y_\lambda^k \mid Y_1^{\lambda-1})}{p_0(Y_\lambda^k \mid Y_1^{\lambda-1})}\,.$$

Hence for the CUSUM and the average likelihood ratio statistics we have

$$U_n =: \max_{\lambda \le n} \log g_\lambda^n = \max(0, \max_{\lambda \le n-1} \log g_\lambda^{n-1}) + \log g_n^n = \log g_n^n + U_{n-1}^+, \quad U_0 = 0, \tag{6.22}$$

$$R_n =: \sum_{\lambda=1}^{n} g_\lambda^n = g_n^n \sum_{\lambda=1}^{n-1} g_\lambda^{n-1} + g_n^n = g_n^n(1 + R_n), \quad R_0 = 0. \tag{6.23}$$

Particularly, in the i.i.d. case $\log g_n^n(Y_1^n) = \Lambda_n(Y_n)$ and the relations (6.22) and (6.23) coincide with (6.2) and (6.3), respectively. In the latter case both statistics are also Markov processes, which helps to evaluate performance of the detection algorithms. But in general the statistics are non-Markov and the analysis is more difficult.

The most important question is how to choose the thresholds in order to guarantee a specified false alarm rate. To answer this question we first observe that $\boldsymbol{E}_0(g_n^n \mid Y_1^{n-1}) = 1$ for any $n \ge 1$ and any model. As a result, it is easily seen that for every $n \ge 1$ the statistic $R_n - n$ is a $\boldsymbol{P}_0$−martingale with mean zero,

$$\boldsymbol{E}_0(R_n - n \mid Y_1^{n-1}) = R_{n-1} - (n-1), \quad \boldsymbol{E}_0 R_n = n,$$

and this is valid for an arbitrary model.

Then, by the optional stopping theorem, for any Markov time τ, $\boldsymbol{E}_0 R_\tau = \boldsymbol{E}_0\tau$, while by the definition of the stopping time $\tilde{\tau}_B$, $\boldsymbol{E}_0 R_{\tilde{\tau}_B} \ge B$. This implies

$$\boldsymbol{E}_0\tilde{\tau}_B \ge B, \tag{6.24}$$

which holds for any model of signal and noise. Therefore if we set

$$B = 1/\overline{\mathrm{Fr}}, \tag{6.25}$$

then the frequency of false alarms will be upper bounded by $\overline{\mathrm{Fr}}$.

It turns out that the martingale property of the statistic R_n can be used to obtain the upper bound for the frequency of false alarms in the CUSUM procedure. Indeed, for any $n \geq 1$

$$\exp(U_n) = \max_{1\leq\lambda\leq n} g_\lambda^n \leq R_n = \sum_{\lambda=1}^{n} g_\lambda^n$$

and hence on the event $\{\tau_a < \infty\}$, $\exp(U_{\tau_a}) \leq R_{\tau_a}$. Thus, if $\boldsymbol{E}_0\tau_a < \infty$, then

$$e^a \leq \boldsymbol{E}_0 \exp(U_{\tau_a}) \leq \boldsymbol{E}_0 R_{\tau_a} = \boldsymbol{E}_0\tau_a\,,$$

where the left inequality follows from the definition of the stopping time τ_a and the right equality from the martingale property of R_n. This shows that for any model (not necessarily i.i.d.) that guarantees finiteness of $\boldsymbol{E}_0\tau_a$ we have the inequality

$$\boldsymbol{E}_0\tau_a \geq e^a\,. \tag{6.26}$$

Hence if

$$a = \log(1/\overline{\mathrm{Fr}}), \tag{6.27}$$

then the frequency of false alarms in the CUSUM-type procedure is upper bounded by $\overline{\mathrm{Fr}}$.

Finally, we note that (6.24) and (6.26) also hold for the adaptive sequential algorithms (6.20), (6.21), since these algorithms are particular cases of (6.22), (6.23) with $g_n^n = \exp\{\Lambda_n^*\}$.

6.7.2. ***The Case of i.i.d. Observations.*** We first observe that in the i.i.d. case the statistic U_n is a Markov process with the transition probability densities (under H_i, $i = 0, 1$),

$$p_i(u_n|u_{n-1}) = f_i\big(u_n - g(u_{n-1})\big), \tag{6.28}$$

where $f_i(y)$ is the probability density of the LLR $\Lambda_k = U_k - g(U_{k-1})$ under H_i; $g(y) = y^+$ for Algorithm 1 (fixed sliding window) and $g(y) = y\mathbb{1}_{\{0<y<a\}}$ for Algorithm 2 with immediate renewal (see (6.14)). So is the statistic $r_n = \log R_n$ with the transition probability density

$$p_i(u_n \mid r_{n-1} = u_{n-1}) = f_i\Big(u_n - \log\big(1 + e^{u_{n-1}}\mathbb{1}_{\{u_{n-1}<b\}}\big)\Big)$$

for the quasi-Bayesian procedure (see (6.17)). The analysis of these algorithms may be done by using the renewal theory (see, e.g., [20, 46, 59]).

In what follows we consider only the CUSUM-type procedure. For the quasi-Bayesian algorithm all formulas are similar. The false alarm probability is

$$P_{\mathrm{fa}}(T, a) = \boldsymbol{P}_0(\tau_a \leq T) = 1 - \boldsymbol{P}_0(U_1 < a, \ldots, U_T < a), \tag{6.29}$$

while the target missing probability is

$$P_{\mathrm{mis}}(T, \gamma, \lambda) = \boldsymbol{P}_{\lambda,\gamma}(\tau_a > T) = \boldsymbol{P}_{\lambda,\gamma}(U_1 < a, \ldots, U_T < a), \tag{6.30}$$

where the statistic U_n obeys the recursion

$$U_n = \Lambda_n + g(U_{n-1}), \quad n = 1, \ldots, T\,. \tag{6.31}$$

The threshold is chosen from the equation

$$P_{\mathrm{fa}}(T, a) = \alpha_0,$$

where α_0 is a specified false alarm rate. Therefore to choose the thresholds and to evaluate the probabilities of correct detection one needs to find the probability of entry in the set $[a,\infty)$ by the Markov process U_n during the time T. This process starts from 0 for Algorithm 1 and from some random point in $(-\infty,a)$ for Algorithm 2.

In general, let $\{X_n, 0 \le n \le T\}$, $X_0 = x$ (x is fixed) be a Markov process with the transition density $f(y,z)$ ($X_{n-1} = z \to X_n = y$). Define the system of functions $p_k(y,x)$ by the recursion

$$\begin{aligned} p_{k+1}(y,x) &= \int_{-\infty}^{a} p_k(y,z) f(z,x)\, dz, \qquad k \ge 1\,, \\ p_1(y,x) &= F(y,x) \qquad \text{where} \quad F(y,x) = \int_{-\infty}^{y} f(z,x)\, dz\,. \end{aligned} \tag{6.32}$$

The function $p_k(y,x)$ is nothing but the conditional probability of the event $\{X_k < y, X_{k-1} < a, \dots, X_1 < a\}$ given $X_0 = x$. Then

$$\boldsymbol{P}(X_1 < a, \dots, X_T < a \mid X_0 = x) = p_T(a,x)\,. \tag{6.33}$$

6.7.3. ***Performance of Algorithm 1.*** Similar to (6.32) for $i = 0,1$, define the functions $p_k^{(i)}(y,x)$ by

$$\begin{aligned} p_{k+1}^{(i)}(y,x) &= \int_{-\infty}^{a} p_k^{(i)}(y,z) f_i(z - x^+)\, dz, \qquad k \ge 1\,, \\ p_1^{(i)}(y,x) &= F_i(y - x^+)\,, \end{aligned} \tag{6.34}$$

where $f_i(y - x^+)$ is the transition density function of the statistic U_n (under H_i) that satisfies (6.31) with $g(U_{n-1}) = U_{n-1}^+$ and the null initial condition, $U_0 = 0$. The probabilities of errors are then computed as follows

$$P_{\text{fa}}(T,a) = 1 - p_T^{(0)}(a,0)\,; \tag{6.35}$$

$$P_{\text{mis}}(T,\lambda,\gamma,a) = p_T^{(1)}(a,0), \quad \text{if} \quad \lambda \le n - T + 1,\ \gamma \ge n\,. \tag{6.36}$$

If $\lambda > n - T + 1$ or/and $\gamma \le n$, then the situation is more delicate. It may be shown that in this case the probability of target missing

$$\begin{aligned} P_{\text{mis}}(T,\lambda,\gamma,a) = \int_{-\infty}^{a} \int_{-\infty}^{a} & p_{n-\gamma}^{(0)}(a,\xi) p_{\gamma-\lambda+1}^{(1)}(d\xi,\eta) p_{\lambda-1}^{(0)}(d\eta,0) \\ \text{if} \quad & \lambda > n - T + 1,\ \gamma < n\,. \end{aligned} \tag{6.37}$$

To obtain the mis-detection probability for $\{\lambda \le n - T + 1, \gamma < n\}$ and $\{\lambda > n - T + 1, \gamma \ge n\}$, one has to put

$$\begin{aligned} p_{\lambda-1}^{(0)}(d\eta,0) &= \delta(\eta)\, d\eta \qquad \text{for} \quad \{\lambda \le n - T + 1, \gamma < n\}\,, \\ p_{n-\gamma}^{(0)}(a,\xi) &= 1 \qquad \text{for} \quad \{\lambda > n - T + 1, \gamma \ge n\} \end{aligned}$$

in (6.37) ($\delta(x)$ being a delta-function). Particularly, this yields the relation (6.36).

It is worth mentioning that the probabilities $p_k^{(i)}(a,0)$ satisfy the integral equations

$$p_k^{(i)}(a,0) = \int_{-\infty}^{a} p_{k-1}^{(i)}(a,z) f_i(z)\, dz, \qquad k \geq 1\,, \tag{6.38}$$
$$p_0^{(i)}(y,x) = 1\,,$$

which are simpler than (6.34).

The formulas (6.35)-(6.37) determine the performance in terms of the probabilities of false alarm P_{fa} and target missing P_{mis} at each current time n within the window of the length T. In particular, the threshold a is found from the equation (see (6.29))

$$1 - \alpha_0 = p_T^{(0)}(a,0)$$

where the function $p_T^{(0)}(a,0)$ satisfies the recursive relation (integral equation) (6.38). Also, the probability of target missing in the interval of its presence, $[\lambda, \gamma]$, can be approximated by[5]

$$P_{\text{mis}}(\gamma - \lambda + 1, a) \approx p_{\gamma-\lambda+1}^{(1)}(a,0)\,.$$

The following elementary estimates are useful for initial choice of threshold and target missing probability evaluation:

$$P_{\text{fa}}(T,a) \geq \boldsymbol{P}_0(S_T \geq a), \quad P_{\text{mis}}(\gamma - \lambda + 1, a) \leq 1 - \boldsymbol{P}_{\lambda,\gamma}(S_\gamma - S_{\lambda-1} \geq a) \tag{6.39}$$

where $S_n = \sum_{k=1}^{n} \Lambda_k$.

Along with the probabilities of errors it is interesting to evaluate the mean time delay in target detection $\boldsymbol{E}_{\lambda,\gamma}\{\tau_a(T,n) - \lambda | \tau_a(T,n) > \lambda\}$ (in order to estimate benefits of the sequential scheme (6.10) compared to the direct non-sequential test (6.8)). Here $\boldsymbol{E}_{\lambda,\gamma}$ is the symbol of expectation for the given values of λ and γ. By

$$I = \int \log\left[\frac{p_1(y)}{p_0(y)}\right] p_1(y)\, dy$$

denote the Kullback-Leibler information number. In case when $T \gg a/I$, $\gamma - \lambda > a/I$, and a is sufficiently large (α_0 is small) the mean detection delay may be approximated by

$$\boldsymbol{E}_{\lambda,\gamma}\{\tau_a(T,n) - \lambda | \tau_a(T,n) > \lambda\} \approx a/I.$$

6.7.4. ***Performance of Algorithm 2.*** For Algorithm 2, reasonable performance may be expressed in terms of the frequency of false alarms $\text{Fr} = 1/\boldsymbol{E}_0\tau_a$ ($\boldsymbol{E}_0\tau_a$ is the mean time between false alarms) and the detection delay $\boldsymbol{E}_{\lambda,\gamma}\{\tau_a - \lambda | \tau_a > \lambda\}$. However, such characteristics as the probabilities of false alarm and target detection in the fixed interval of the length T are also of interest.

For Algorithm 2 with immediate renewal we have

$$\text{Fr} \leq \exp(-a), \quad \boldsymbol{E}_{\lambda,\gamma}\{\tau_a - \lambda | \tau_a > \lambda\} \approx a/I. \tag{6.40}$$

[5]This formula is approximate, since $U_{\lambda-1}$ is a random variable belonging to the interval $(-\infty, a)$. For $\lambda = 1$ it is exact.

The first formula in (6.40) follows from (6.26) and the second formula is valid when $\gamma - \lambda > a/I$ and a is sufficiently large (Fr is small). According to the first inequality in (6.40) the threshold may be chosen as $a = \log(1/\overline{\mathrm{Fr}})$ to guarantee the frequency of false alarms $\mathrm{Fr} \le \overline{\mathrm{Fr}}$, where $\overline{\mathrm{Fr}}$ is a specified level.

However, it is even more reasonable to choose the threshold based on the specified probability of false alarm $P_{\mathrm{fa}}(T, a)$ within the time interval of the length T at each current point in time n. Then the mean detection delay may be approximately evaluated by the second equality in (6.40) and the algorithm is optimal – it minimizes the mean detection delay among all algorithms with this kind of constraints (at least for sufficiently large a).

To obtain the probabilities of errors in the fixed interval T we note that the statistic U_n is again the Markov process which obeys the recursion (6.31) with the random initial condition $U_0 = u$ and with $g(U_{n-1}) = U_{n-1}\mathbb{1}_{\{0<U_{n-1}<a\}}$. The random variable u has the distribution $G_0(y)$, which is nothing but a stationary distribution of the Markov process $\{U_n\}$ under the hypothesis H_0 (target is absent). This distribution satisfies the integral equation

$$G_0(x) = \int F_0(x - s\mathbb{1}_{\{0<s<a\}})\, dG_0(s),$$

where $F_0(y)$ is the distribution function of Λ_n under H_0. Thus, the false alarm probability is found as

$$P_{\mathrm{fa}}(T, a) = \int_{-\infty}^{a} P_{\mathrm{fa}}(T, a, x)\, dG_0(x),$$

where $P_{\mathrm{fa}}(T, a, x)$ is the probability (6.29) in case when $U_0 = x$, $x < a$. This last probability is computed according to (6.33) with $p_T(a, x) = p_T^{(0)}(a, x)$, which is defined by the recursive relation

$$p_k^{(0)}(a, x) = \int_{-\infty}^{a} p_{k-1}^{(0)}(a, z) f_0\big(z - x\mathbb{1}_{\{0<x<a\}}\big)\, dz, \quad k \ge 1\,, \qquad p_0^{(0)}(a, x) = 1\,.$$

As we mentioned above, the analysis of the proposed algorithms may be done based on the renewal theory. This theory allows for more accurate performance evaluation as compared to the above formulas (for the mean detection delay and frequency of false alarms). The detailed analysis of the proposed algorithms and their tuning will be done in the near future (see also Section 6.8, relationships (6.41),(6.42) for more accurate estimates in a Gaussian case).

6.8. **TbD Model with Spatio-Temporal Matched Filter.** Let us apply the above results to the case where the target moves along a deterministic trajectory. Then the spatio-temporal matched filter is the optimal tool for TbD. Since target's velocity and direction of motion are unknown, the bank of filters should be used and the detection is done in each channel independently. Assume also that after clutter suppression the residual clutter (plus noise) samples are i.i.d. and Gaussian, $\mathcal{N}(0, \sigma^2)$ (see the model (5.9)). Then the LLR Λ_k is

$$\Lambda_k = \frac{1}{\sigma^2} \sum_{i,j} h(x - x_{ij} - (k-1)v\Delta) Z_{ij}(k) - \frac{1}{2\sigma^2} \sum_{i,j} h(x - x_{ij} - (k-1)v\Delta)^2\,.$$

In the above formula we implicitly used the assumption that the target moves with the constant given speed v and in a given direction (i.e., in the particular "velocity-angle" channel). This

formula is readily generalized for arbitrary (but known) dynamics. Note that here we use the notation $Y_k = \{Z_{ij}(k)\}$.

According to (6.31) the statistic $U_n(x)$ is calculated as

$$U_n(x) = \Lambda_n(x) + g(U_{n-1}(x)), \quad n = 1, 2, \dots$$

Then the above algorithms are applied in each channel independently (here we mean the channels that are related to different values of the initial position x).

If the signal is exactly located in the given channel, then the LLRs $\{\Lambda_k(x)\}$ are i.i.d. Gaussian random variables with parameters

$$\boldsymbol{E}_1\Lambda_k = -\boldsymbol{E}_0\Lambda_k = \tfrac{1}{2}\rho, \quad \mathrm{Var}_0\Lambda_k = \mathrm{Var}_1\Lambda_k = \tfrac{1}{4}\rho,$$

where $\rho = \frac{1}{\sigma^2}\sum_{i,j} h(x_{ij})^2$ is SNR.

Performance of the developed algorithms is evaluated by formulas obtained above in Section 6.7 with the densities

$$f_0(y) = \frac{1}{\sqrt{\rho}}\varphi\left(\frac{2y+\rho}{\sqrt{\rho}}\right), \quad f_1(y) = \frac{1}{\sqrt{\rho}}\varphi\left(\frac{2y-\rho}{\sqrt{\rho}}\right),$$

where $\varphi(y)$ is a standard normal density function,

$$\varphi(y) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\}.$$

In particular, for Algorithm 1 (sliding window), the simplest estimates for probabilities of errors are given by (see (6.39))

$$P_{\mathrm{fa}}(T, a) \geq 1 - \Phi\left(\frac{2a+\rho T}{\sqrt{\rho T}}\right), \quad P_{\mathrm{mis}}(\gamma - \lambda + 1, a) \leq \Phi\left(\frac{2a - \rho(\gamma - \lambda + 1)}{\sqrt{\rho(\gamma - \lambda + 1)}}\right)$$

where $\Phi(y)$ is a standard normal distribution function.

Also, for the sequential Algorithm 2 with immediate renewal for sufficiently large a and $\gamma > 2a/\rho$ the following approximate estimates hold:

$$\begin{gathered} \mathrm{Fr} \lessapprox \beta e^{-a}, \\ \boldsymbol{E}_{\lambda,\gamma}\{\tau_a - \lambda \mid \tau_a \geq \lambda\} \approx \tfrac{2}{\rho}(a + \varkappa + C), \end{gathered} \tag{6.41}$$

where

$$\begin{cases} \beta &= \frac{2}{\rho}\exp\left\{-2\sum_{k=1}^{\infty}\frac{1}{k}\Phi\left(-\frac{1}{2}\sqrt{\rho k}\right)\right\}; \\ \varkappa &= 1 + \frac{\rho}{4} - \sqrt{\rho}\sum_{k=1}^{\infty}\left[\frac{1}{\sqrt{k}}\varphi\left(\frac{1}{2}\sqrt{\rho k}\right) - \frac{1}{2}\sqrt{\rho k}\,\Phi\left(-\frac{1}{2}\sqrt{\rho k}\right)\right]; \\ C &= \boldsymbol{E}_1\left[\min\left\{0, \min_{n\geq 1}\sum_{k=1}^{n}\Lambda_k\right\}\right] \in [0, -1]. \end{cases} \tag{6.42}$$

These formulas were obtained by using the nonlinear renewal theory [59]. They improve the upper bounds for the frequency of false alarms and the estimates for the mean detection delay presented above in the general case (see (6.26), (6.27), (6.40)).

Instead of making decisions on target presence in each channel we may apply the algorithm, which is based on the maximum likelihood statistic,

$$M_n = \max_x U_n(x), \quad n = 1, 2, \dots$$

This statistic, however, does not allow for a convenient recursive computation. To simplify the algorithm we use the "trick", which was already used in Section 6.6 to build adaptive detection algorithms in general case. By $x_n^* = x_n^*(\boldsymbol{Z}^n)$ denote the maximum likelihood estimate of the target location, i.e.

$$x_n^* = \arg\max_x \sum_{k=1}^{n} \Lambda_k(x).$$

The decision statistic is defined by the recursive formula

$$U_n^* = \Lambda_n(x_{n-1}^*) + g(U_{n-1}^*), \quad n = 2, 3, \dots, \quad U_1^* = 0.$$

The "stopping rule" is then defined by

$$\tau_a^* = \inf\{n \geq 2 : U_n^* \geq a\}.$$

Thus, we use the one-stage delayed estimates of the signal location in the partial LLR Λ_n. The threshold is chosen according to the relation (6.27), which allows us to upper bound the false alarm rate. This algorithm is especially attractive when only a single target is expected. In the multi-target situation the previous algorithm (decisions in independent channels) is perhaps the best one can do.

6.9. **TbD Model with Optimal Nonlinear Filter.** In the context of TbD with optimal nonlinear filtering (see Section 5) the estimate of the partial LLR Λ_n obtained on the basis of previous $n-1$ observations is defined by

$$\Lambda_n^* = \sum_{i,j} \hat{h}_{n-1}(x_{ij}) Z_{ij}(k) - \frac{1}{2} \sum_{i,j} \hat{h}_{n-1}(x_{ij})^2, \tag{6.43}$$

where $\hat{h}_k(x_{ij}) = h(\hat{X}_k - x_{ij})$ and $\hat{X}_k$ is the estimate of the signal location based on the k observations. The latter estimate is formed at the output of the optimal nonlinear filter that tracks the target location. We will consider two estimates – the mean estimate and the maximum *a posteriori* estimate, which are defined below in (7.3) and (7.4) (see Section 7). Thus, the above results are applied by using the LLR estimate (6.43).

To be precise, the adaptive CUSUM statistic is defined by the recursion

$$U_n^* = \Lambda_n^* + g(U_{n-1}^*), \quad n = 2, 3, \dots, \quad U_1^* = 0 \tag{6.44}$$

and the "stopping rule" is

$$\tau_a^* = \inf\{n \geq 2 : U_n^* \geq a\}. \tag{6.45}$$

In other words, again we have used the same trick – the true LLR Λ_n at stage n is replaced with its estimate $\Lambda_n^* = \Lambda_n(\hat{X}_{n-1})$ obtained based on the previous $n-1$ observations. Instead of $\hat{X}_{n-1}$ one may use the one step predicted (extrapolated) estimate X_n^*.

The alternative algorithm (see (6.19), (6.21)) is of the form

$$\tilde{\tau}_B^* = \inf\{n \geq 2 : R_n^* \geq B\}, \tag{6.46}$$

where the estimate of the average likelihood ratio R_n^* satisfies the recursion

$$R_n^* = \exp\{\Lambda_n^*\}(1 + R_{n-1}^* \mathbb{1}_{\{R_{n-1}^* < B\}}), \quad n = 2, 3, \ldots, \quad R_1^* = 0\,. \tag{6.47}$$

The results of application of these algorithms will be discussed in Section 7.4.

7. Testing of the Developed Algorithms for IRST Data. Results of Experiments and Simulation

In the examples considered below because of the uncertainty of possible behavior of a non-cooperative target, the alignment of successive frames is extremely difficult. The frame alignment is done recursively along the maximum posterior distribution path on the basis of the optimal nonlinear filtering. Use of nonlinear algorithms is necessitated by the specifics of the observation structure. Indeed, in TbD the observation function (signal) $S(k, \boldsymbol{r})$ is highly sharp (essentially a delta-function in small domain like 2×2 or 3×3 pixels). In addition, IR imaging sensors provide angle-only measurements. These types of measurements, especially in low S(N+C)R situations, practically rule out application of extended Kalman and similar filters. As already mentioned above, application of multi-dimensional matched filters and banks of velocity filters is possible only for constant speed targets with linear trajectories, which is not the case in the examples considered. At the same time the proposed ONF-based algorithm works perfectly well even for very low S(N+C)R (up to -6dB after clutter removal). Particularly, it may be seen from the figures presented below that the uncertainty is completely overcome after processing of several frames (in these examples S(N+C)R ranges from -1dB to -7dB after simple pre-processing and clutter removal).

The developed adaptive detection algorithms that use the estimates of target location constructed on the basis of ONF TbD work also perfectly well. Tracks of randomly appearing and disappearing targets are reliably detected up to -7dB and the algorithms allow us to obtain low false alarm rate even in a heavy cluttered background.

7.1. **Clutter Removal: Real IRST Data.** The proposed nonparametric approach to clutter removal (see Section 4.1) has certain advantages over conventional methods. This fact is illustrated by Figure 7 which presents the results of kernel smoothing of IR data. The 'local mean removal' procedure used for clutter rejection in a series of papers by Reed et al. (see [39], [38]) is equivalent to a kernel estimator with a product kernel K generated by uniform kernels $K_1(x) = K_2(x) = 0.5 I_{\{|x| \leq 1\}}$. It may be seen that the Fuller kernel (see [26]) provides superior smoothing performance compared to the uniform kernel. The reason is that our approach relies on the trade-off between a squared bias and a variance of estimators while the uniform kernel minimizes only the asymptotic variance of estimators and hence induces a large bias term.

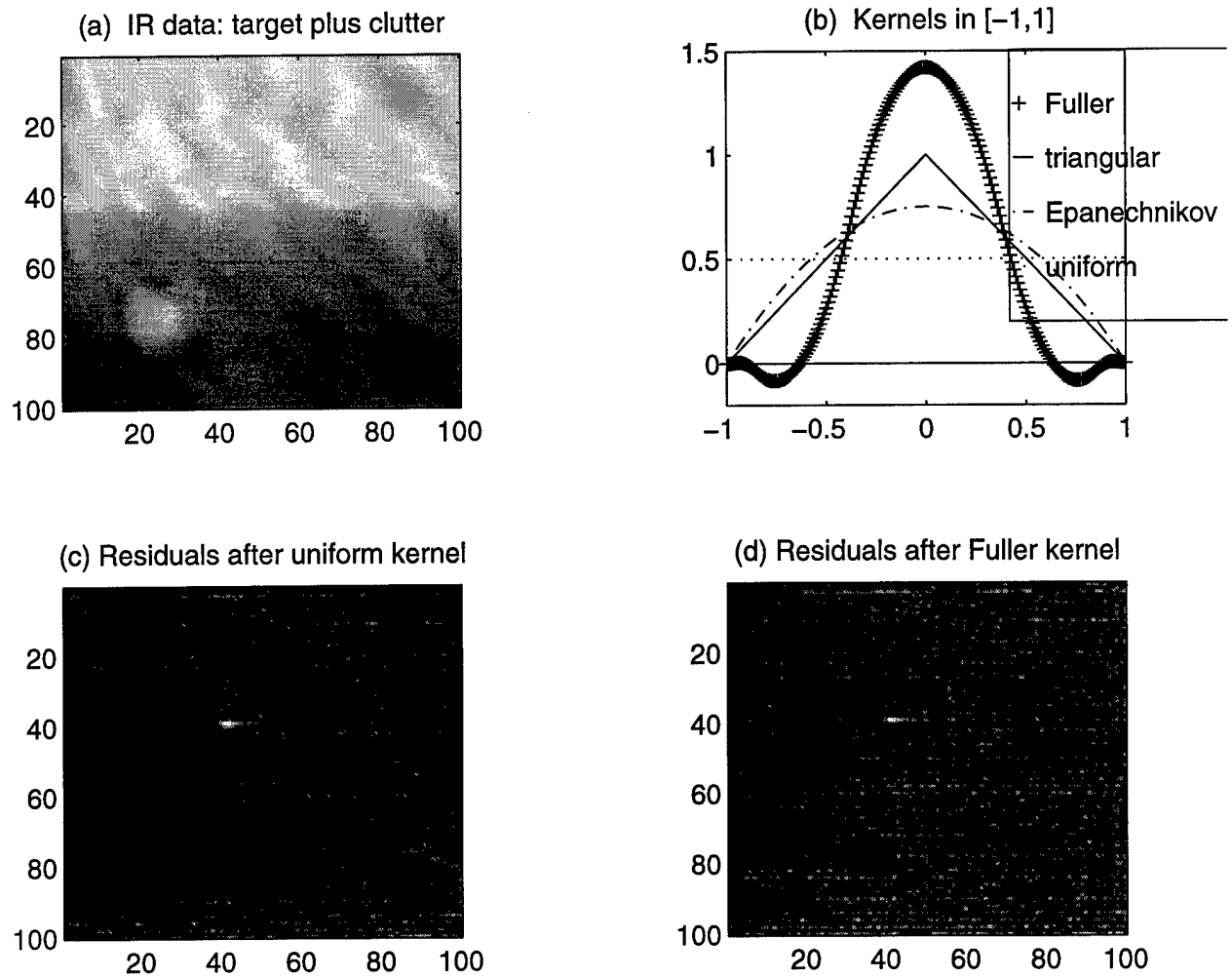


Figure 7. NAVY IRSS sensor - the LAPTEX field test. Original IR image (a) and the results of clutter removal by uniform kernel (c) and Fuller kernel (d)

7.2. **Example 1: TbD of a Target Based on IRST Data.** The first example presented is a naïve model of an actively maneuvering air target which flies overhead over a stationary observer situated on the ground. The complete field of view of the observer is modeled by a square $[0,1] \times [0,1]$ with the horizon line at $y = 0.1$. The target initially appears close to the horizon at a point uniformly distributed inside a rectangle $[0.2, 0.8] \times [0.1, 0.15]$, then the horizontal coordinate evolves as a Wiener process with variance $\alpha^2 t$, and the vertical coordinate as an exponent $e^{\beta t}$.

Formally, the trajectory model is described by a stochastic differential equation[6]

$$X_0 \sim \text{Unif}(0.2, 0.8), \qquad Y_0 \sim \text{Unif}(0.1, 0.15);$$
$$dX_t = \alpha\, dW_t, \quad \alpha = 0.05, \qquad dY_t = \beta Y_t\, dt, \quad \beta = 0.6.$$

The target was assumed to emit a 3×3 pixels signature which was observed with additive Gaussian white noise. More precisely, let δ be the size of a pixel in the observed image in both

[6]Here the components of the vector $\boldsymbol{r}$ are denoted by (x, y). Also $\boldsymbol{X} = (X, Y)$.

horizontal and vertical directions. The target shape is given by

$$h(x,y) = \mathbb{1}_{\{|x|\le 3\delta/2\}}\mathbb{1}_{\{|y|\le 3\delta/2\}}, \quad (x,y)\in\mathbb{R}^2,$$

and the observation at time moment $t_k = k\Delta$, $\Delta = 0.05$ with noise of standard deviation σ and $\text{SNR} = 20\log\frac{1}{\sigma}$ is

$$Z_{ij}(k) = h(X_{k\Delta} - (i-1)\delta, Y_{k\Delta} - (j-1)\delta) + \sigma\xi_{ij}(k), \quad \xi_{ij}(k)\sim \text{Norm}(0,1) \text{ i.i.d.}$$

Examples of a target trajectory and IR data images are shown in Figure 8 and Figure 9.

Several experimental results follow.

- For moderate noise ($\sigma = 1.232$, SNR$= -1.8$dB) the simplest and fastest Haar basis works quite well (see Figure 10).
- For more intense noise ($\sigma = 1.581$, SNR$= -4.0$dB), the algorithm utilizing Haar basis loses track occasionally but later recovers it (see Figure 11).
- For borderline cases ($\sigma = 1.679$, SNR$= -4.5$dB), the Haar basis no longer provides good tracking, however smoother wavelet bases of the same resolution (such as Coiflet-1) still achieve good tracking accuracy (see Figures 12 and 13; both simulations were run on exactly the same observations, and the difference in the quality of tracking is quite obvious).
- With larger noise ($\sigma = 2.109$, SNR$= -6.5$dB), the optimal filter fails to track the target, no matter what basis is utilized.

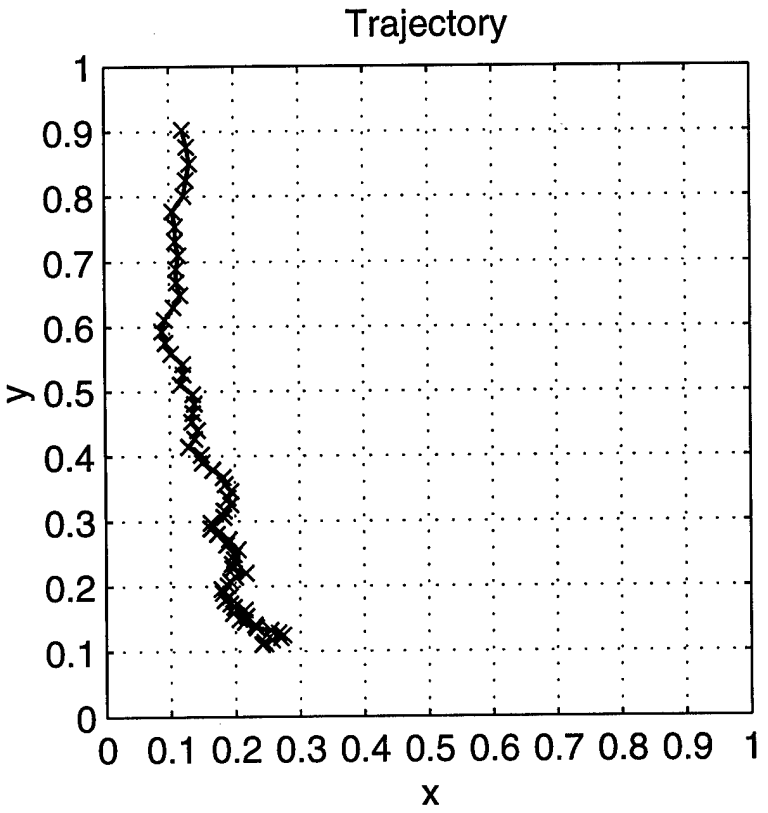


Figure 8. Typical trajectory of overhead target

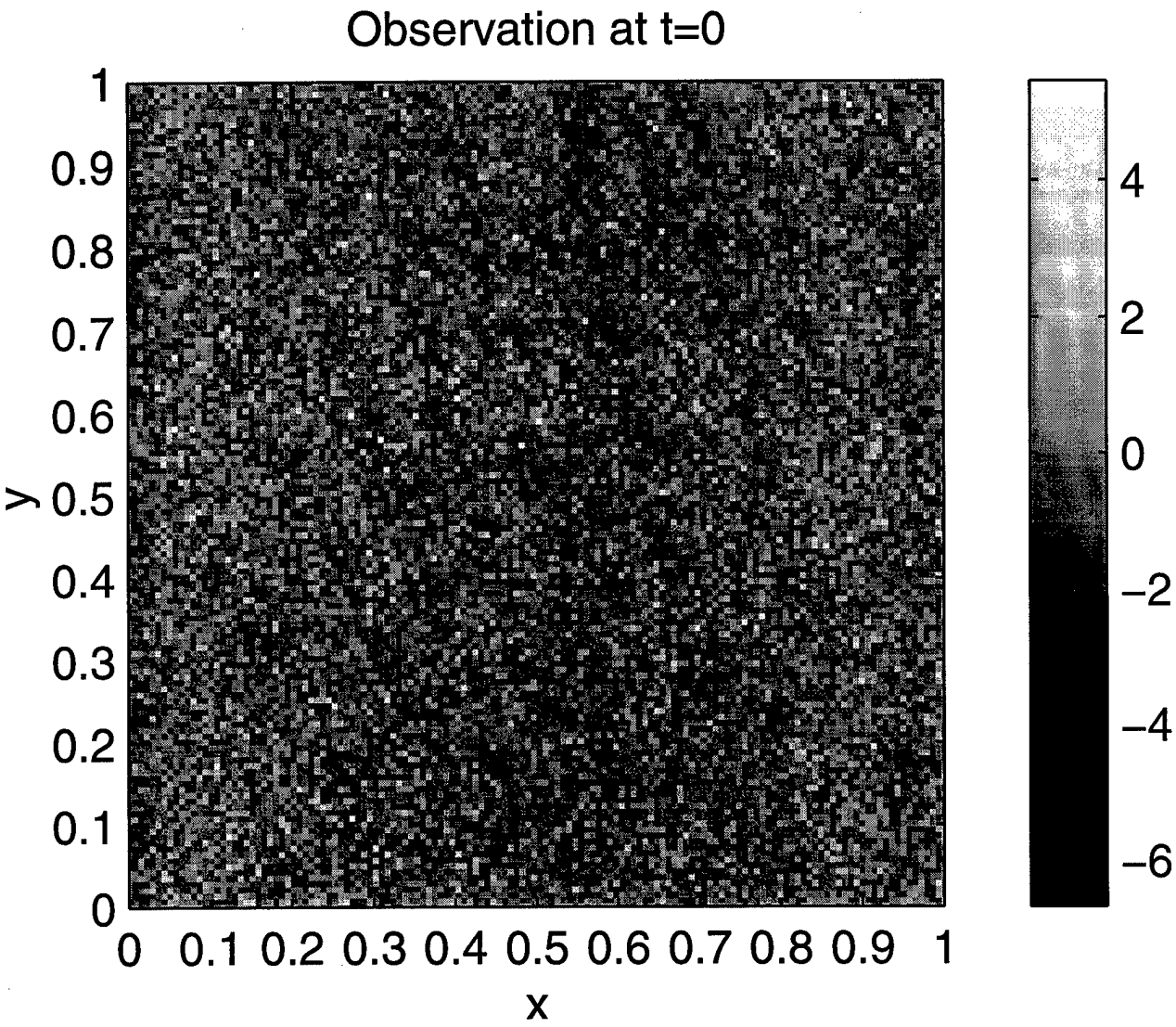


Figure 9. Sample of a single frame of observation

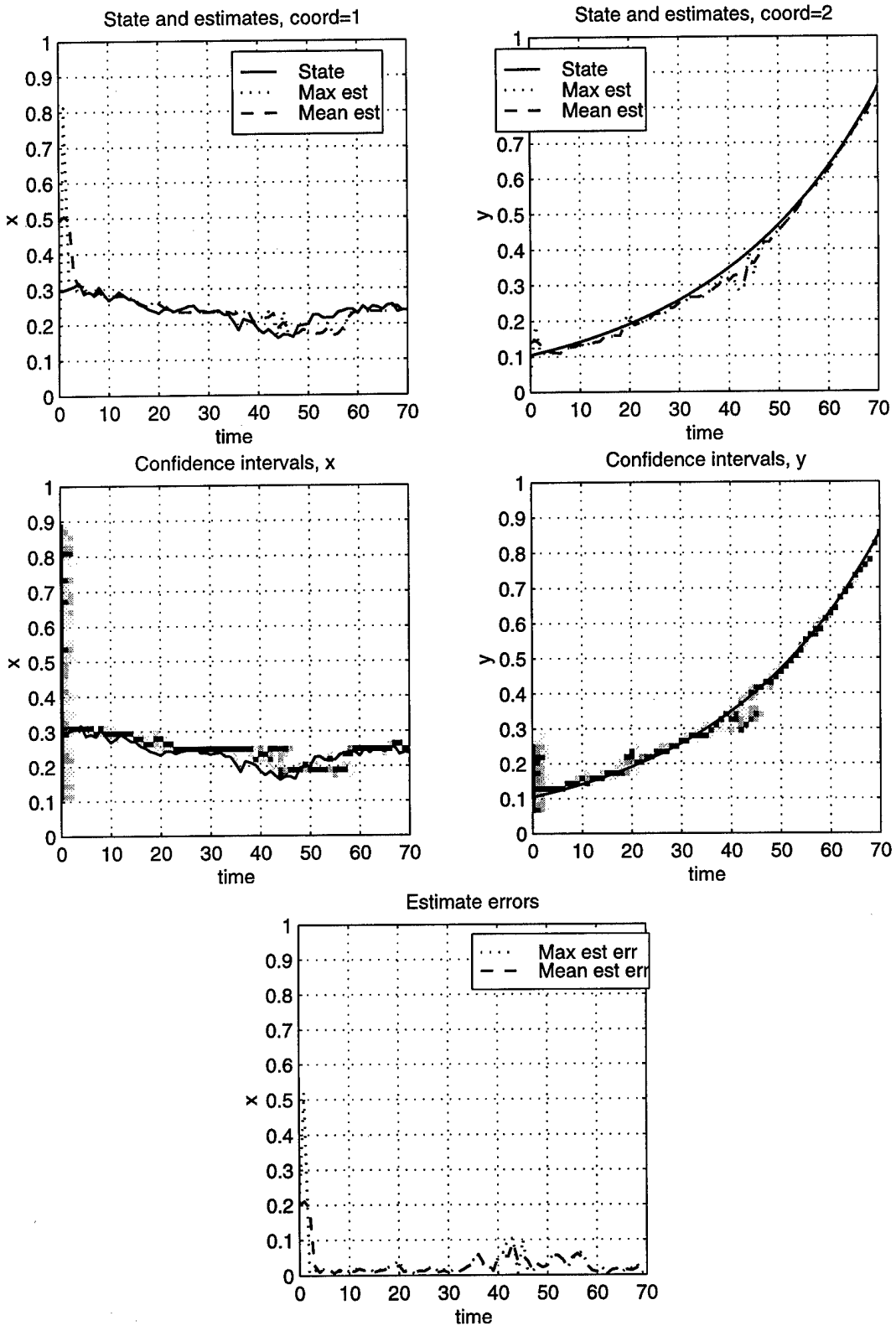


Figure 10. ONF, Haar 64×64 basis, $\sigma = 1.232$, SNR$= -1.8$dB

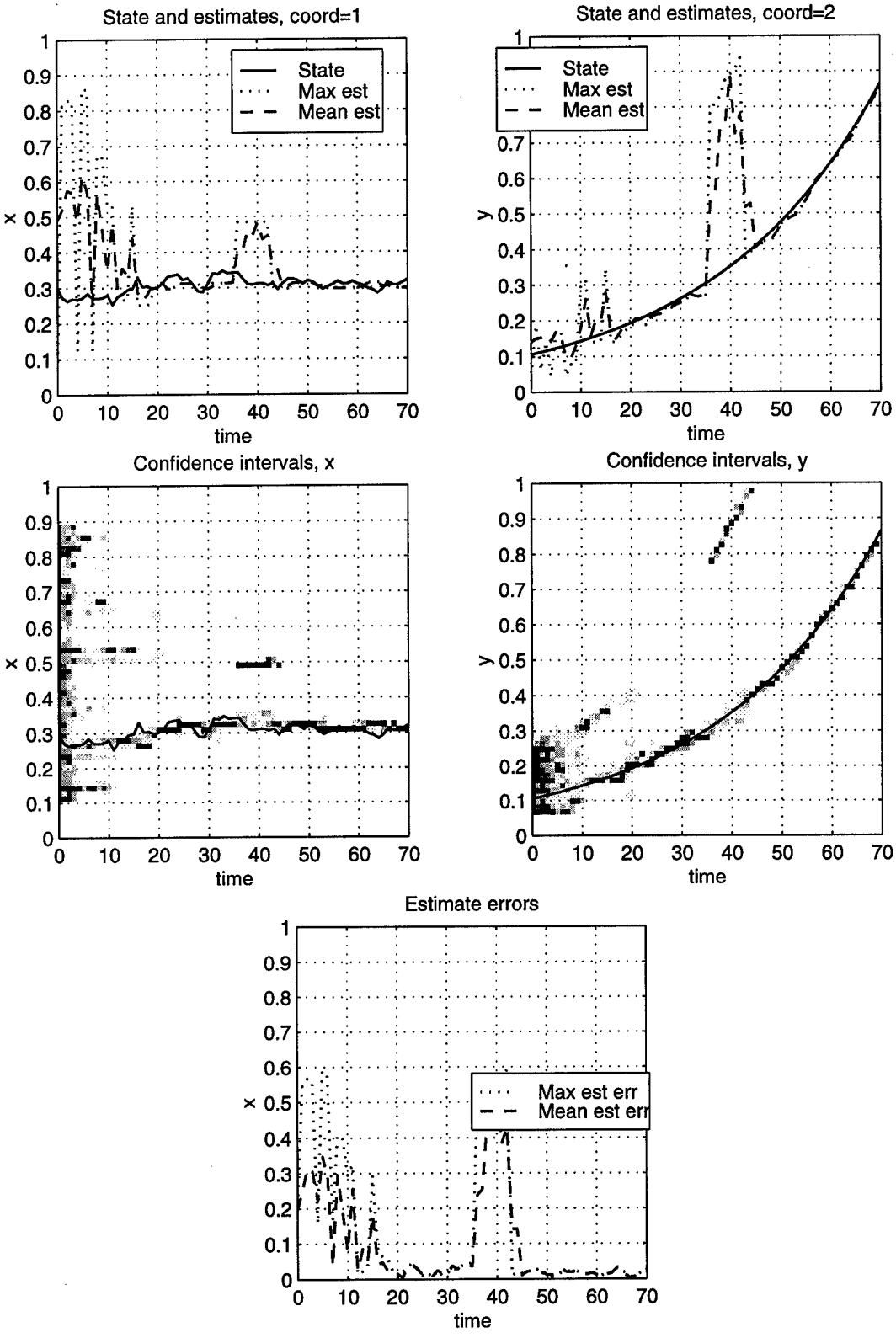


Figure 11. ONF, Haar 64×64 basis, $\sigma = 1.581$, SNR$= -4.0$dB

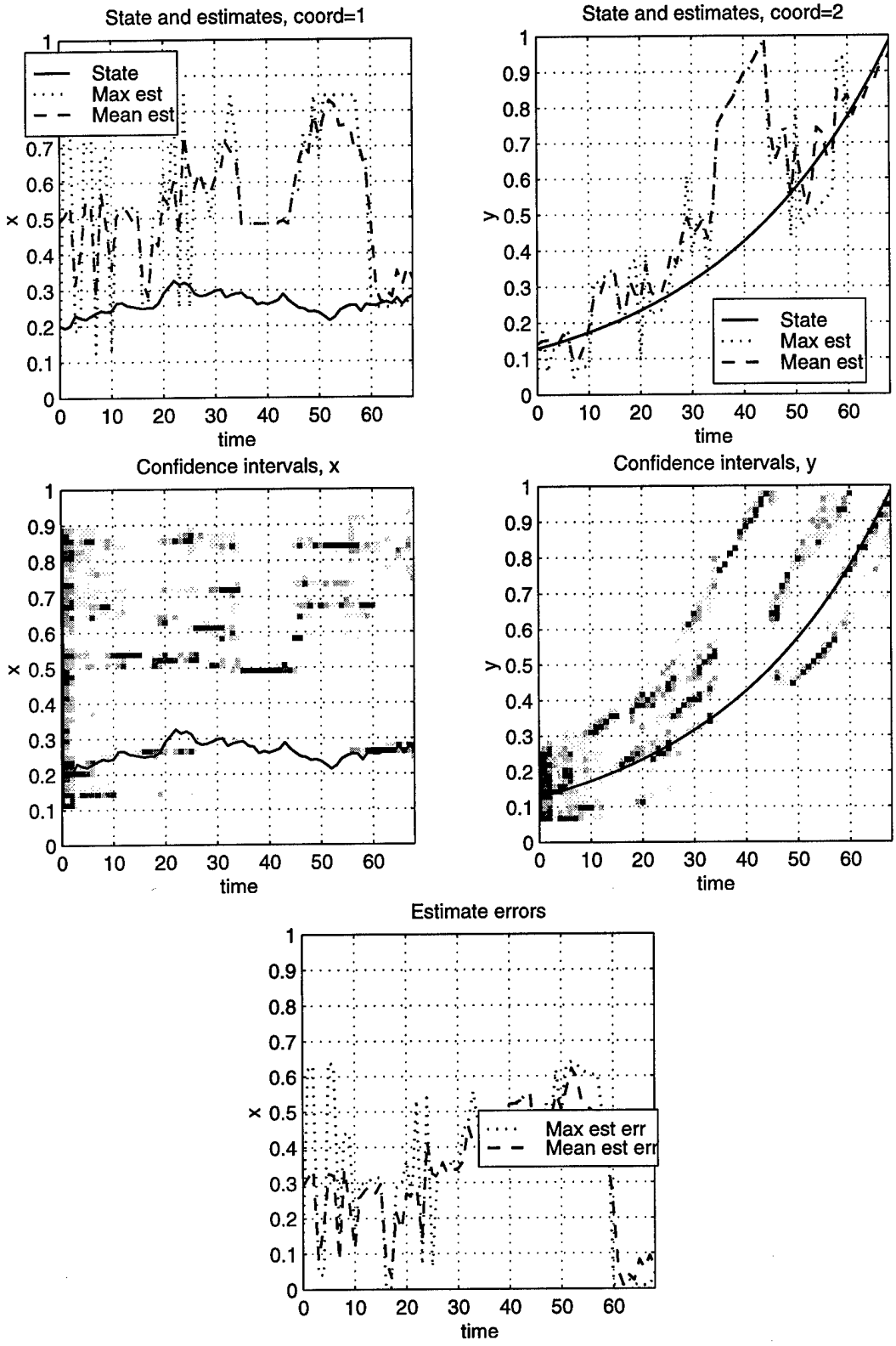


Figure 12. ONF, Haar 64×64 basis, $\sigma = 1.679$, SNR$= -4.5$dB

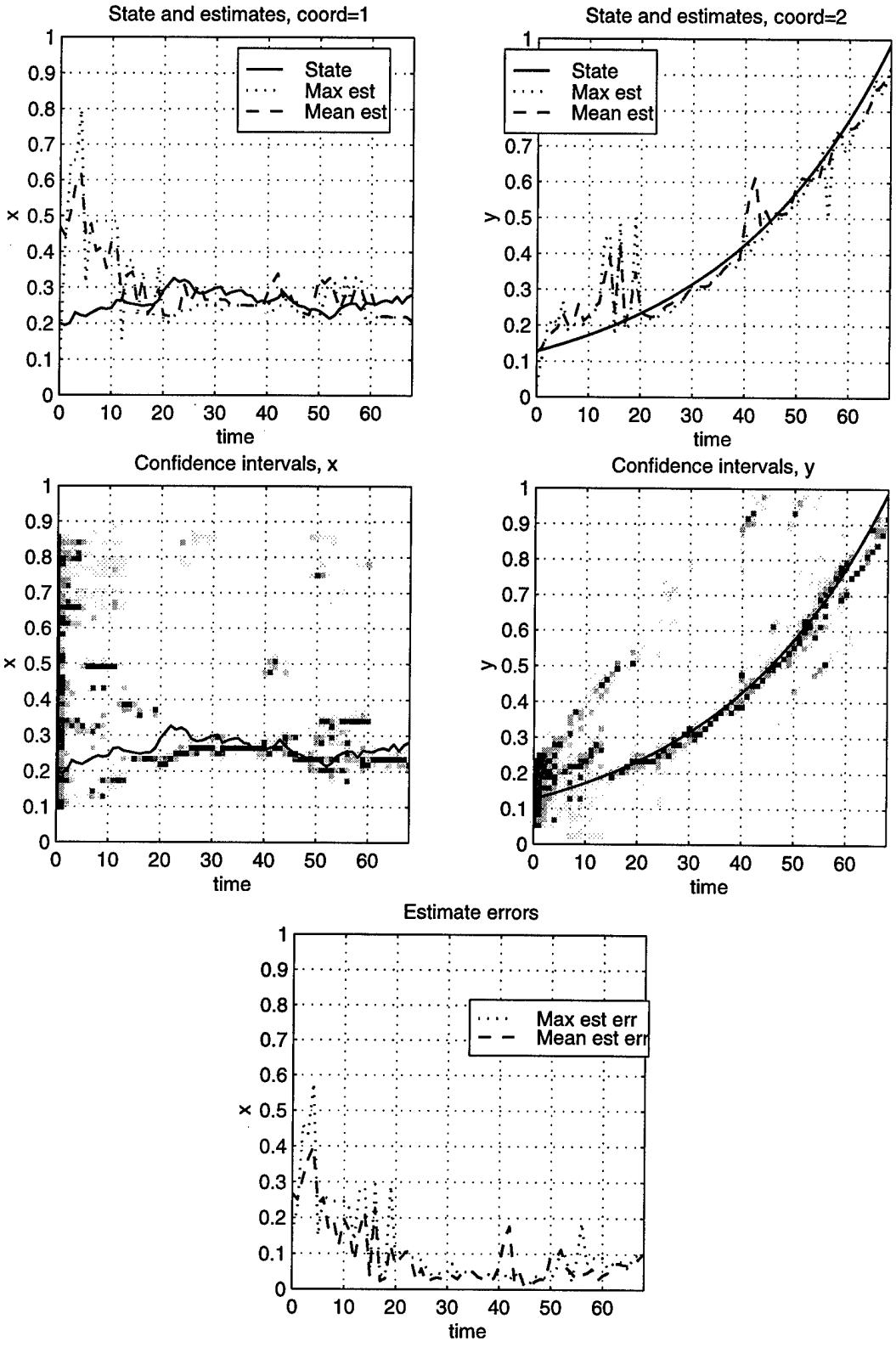


Figure 13. ONF, Coiflet-1 64×64 basis, $\sigma = 1.679$, SNR$= -4.5$dB

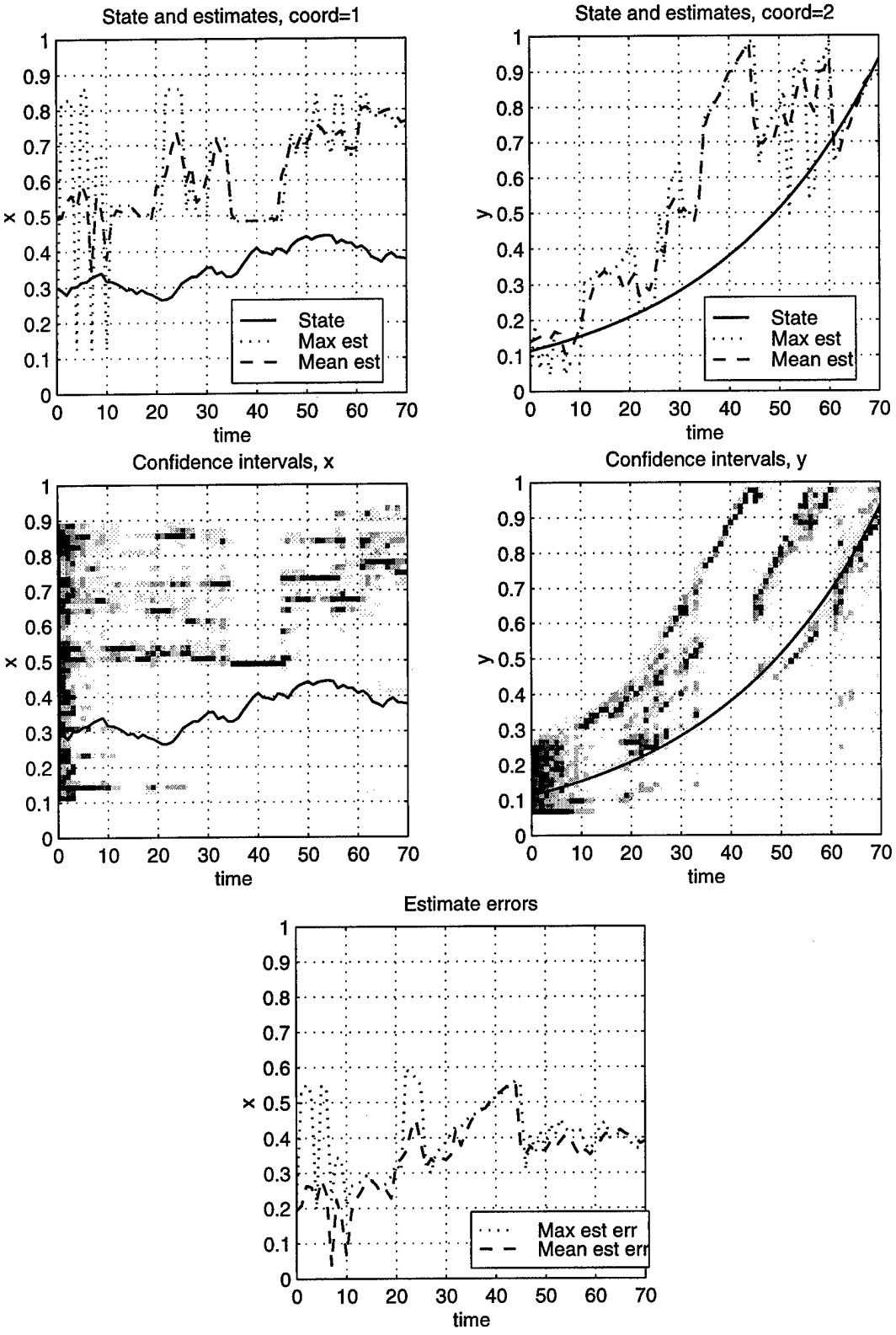


Figure 14. ONF, Haar 64×64 basis, $\sigma = 2.109$, SNR= -6.5dB

7.3. **Example 2: TbD of a Surface Skimming Missile.** Assume that a stationary observation platform (camera) is at a fixed height h_2 above the sea level, and a target approaches with constant velocity v at constant height $h_1 < h_2$ above the sea level. Denote the Earth radius $R = 63,750,000m$. Projection center (the focal point of the camera) is situated at the point with coordinates $(0, 0, R + h_1)$. At time moment $t = 0$ the target appears at the horizon at angle β clockwise from the direction of $y-$axis and moves at angle γ to the line of sight from the observer. Because of rotational invariance after projecting onto the observation cylinder, assume that $\beta = 0$.

The three dimensional coordinate system used next has the origin at the center of the Earth, the $z-$axis passing vertically through the observer, the $y-$axis passing horizontally through the observer in the direction of the point where the target becomes first visible on the horizon, and $x-$axis oriented so that the xyz coordinate system is a Cartesian right hand coordinate system.

$$
\begin{aligned}
x_0 &= 0, \\
y_0 &= \frac{R}{R+h_1}\left(\sqrt{2Rh_1+h_1^2}+\sqrt{2Rh_2+h_2^2}\right), \\
z_0 &= \frac{R^2}{R+h_1}-\frac{1}{R+h_1}\sqrt{2Rh_1+h_1^2}\sqrt{2Rh_2+h_2^2}, \\
x_t &= (R+h_2)\sin\gamma\,\sin\frac{vt}{R+h_2}, \\
y_t &= y_0\cos\frac{vt}{R+h_2}-z_0\cos\gamma\,\sin\frac{vt}{R+h_2}, \\
z_t &= z_0\cos\frac{vt}{R+h_2}+y_0\cos\gamma\,\sin\frac{vt}{R+h_2}.
\end{aligned}
$$

After projecting onto an observation cylinder with radius r (the focal length of the camera), the equation of the motion in two dimensional coordinates (θ, z^p) are as follows

$$
\theta_t = \arctan\frac{x_t}{y_t}, \quad z_t^p = \frac{rz_t}{\sqrt{x_t^2+y_t^2}}.
$$

The observed signal was assumed to be a 3×3 pixels target with additive Gaussian white noise. More precisely, let δ_1 be the size of a pixel in the horizontal (coordinate θ) direction, and δ_2 be the pixel size in the vertical (coordinate z^p) direction. The target shape is given by

$$
h(x_1, x_2) = \mathbb{1}_{\{|x_1|\le 3\delta_1/2\}}\mathbb{1}_{\{|x_2|\le 3\delta_2/2\}}, \quad (x_1, x_2) \in \mathbb{R}^2, \tag{7.1}
$$

and the observation at time moment $t_k = k\Delta$ with noise of standard deviation σ and SNR $= 20\log\frac{1}{\sigma}$ is

$$
\begin{aligned}
\tilde{\theta}_i &= \theta_{min} + (i-1)\delta_1, \\
\tilde{z}_j^p &= z_{min}^p + (j-1)\delta_2, \\
Z_{ij}(k) &= h(\theta_{k\Delta}-\tilde{\theta}_i, z_{k\Delta}^p - \tilde{z}_j^p) + \sigma\xi_{ij}(k), \\
\xi_{ij}(k) &\sim \mathrm{Norm}(0,1) \text{ i.i.d.}
\end{aligned} \tag{7.2}
$$

It is important to emphasize that the Gaussian model for the residual clutter and noise is used only in ONF and detection algorithm development. The algorithms were tested with the use of

the real IR background obtained from SPAWAR. Particularly, the figures in Section 7.4 below contain the results of testing for these real cluttered images (i.e., $\{\xi_{ij}(k)\}_{k\geq 1}$ in (7.2) are given realistic frames) with artificially inserted targets whose motion and shape correspond to models (7.1), (7.2).

7.3.1. ***TbD of Several Targets.*** Ability to detect and track several independent targets was implemented by running the corresponding number of filters in parallel.

The exact updating formulas are given below.

Translation operators. The operator T^i_Δ translates the argument of a function defined on the state space along the trajectory of the model $\#i$. Let us denote the trajectory of the model $\#i$ starting at the point (x,y) at time moment $t=0$ by $q^i(t,x,y)$, and its inverse with respect to the variable t (with variables x and y fixed) by $t^i(q,x,y)$. Let the function to translate be $f:[0,a)\times[0;b]\times\{1\}\cup[0,a)\times[-b,b]\times\{-1\}\cup\{0\}\times\{0\}\times\{0\}\to\mathbb{R}$. Several cases have to be split in order to describe the translation operator:

$$
\begin{aligned}
&\big(T^i_\Delta f\big)(x,y,1)=\tfrac{1}{aq^i(x,0,\Delta)}f(0,0,0),\quad \text{for } x\in[0,a),\quad y\in[0,q^i(x,0,\Delta)),\\
&\big(T^i_\Delta f\big)\big(q^i(\Delta,x,y),1\big)=f(x,y,1),\quad \text{for } x\in[0,a),\quad y\in[0,q^i(x,b,-\Delta)),\\
&\big(T^i_\Delta f\big)\big(q^i(\Delta,x,y),-1\big)=f(x,y,1),\quad \text{for } x\in[0,a),\quad y\in[q^i(x,b,-\Delta),b),\\
&\big(T^i_\Delta f\big)\big(q^i(\Delta,x,y),-1\big)=f(x,y,-1),\quad \text{for } x\in[0,a),\quad y\in[0,b),\\
&\big(T^i_\Delta f\big)(0,0,0)=e^{-\mu t}f(0,0,0)+\int_0^a\int_{-b}^{q^i(x,-b,-\Delta)}f(x,y,-1)dydx.
\end{aligned}
$$

Updating formulas. The unnormalized filtering densities $p^l_t(x,y,u)$ corresponding to the model $\#l$ are recomputed based on observations $\boldsymbol{Z}_k=\{Z_{ij}(k\Delta)\}$ with time step Δ according to formulas

$$
\begin{aligned}
&p^l_0(0,0,0)=1,\\
&p^l_0(x,y,u)=0,\quad \text{for } (x,y,u)\neq(0,0,0),
\end{aligned}
$$

$$
\begin{aligned}
p^l_{k\Delta}(0,0,0)&=\big(T^l_\Delta p^l_{(k-1)\Delta}\big)(0,0,0),\\
H_{k\Delta}(x,y,u,\boldsymbol{Z}_k)&=\exp\Big\{\tfrac{1}{\sigma^2}\sum_{i,j}h(x-x_i,y-y_j,u)Z_{ij}(k\Delta)-\tfrac{1}{2\sigma^2}\sum_{i,j}h(x-x_i,y-y_j,u)^2\Big\},\\
p^l_{k\Delta}(x,y,u)&=H_{k\Delta}(x,y,u,\boldsymbol{Z}_k)\big(T^l_\Delta p^l_{(k-1)\Delta}\big)(x,y,u),\quad \text{for } (x,y,u)\neq(0,0,0).
\end{aligned}
$$

Next, the filtering densities are normalized as follows:

$$
\begin{aligned}
n^l_{k\Delta}&=p^l_{k\Delta}(0,0,0)+\int\int p^l_{k\Delta}(x,y,1)\,dx\,dy+\int\int p^l_{k\Delta}(x,y,-1)\,dx\,dy,\\
\pi^l_{k\Delta}(x,y,u)&=\frac{p^l_{k\Delta}(x,y,u)}{n^l_{k\Delta}},
\end{aligned}
$$

and two estimates of the position of the l-th target are computed. Namely, the mean-square-error estimate:

$$\begin{aligned}\hat{X}^{l,\mathrm{mean}}_{k\Delta} &= \int\int x\pi^l_{k\Delta}(x,y,1)\,dx\,dy + \int\int x\pi^l_{k\Delta}(x,y,-1)\,dx\,dy,\\ \hat{Y}^{l,\mathrm{mean}}_{k\Delta} &= \int\int y\pi^l_{k\Delta}(x,y,1)\,dx\,dy + \int\int y\pi^l_{k\Delta}(x,y,-1)\,dx\,dy,\end{aligned} \tag{7.3}$$

and the maximum *a posteriori* estimate:

$$(\hat{X}^{l,\max}_{k\Delta}, \hat{Y}^{l,\max}_{k\Delta}) = \arg\max_{x,y} \pi^l_{k\Delta}(x,y,u). \tag{7.4}$$

Both estimates can be used in different situations. The first estimate is theoretically optimal. However, it usually takes longer to converge to the real position of the target. The second estimate "detects" the target much faster but is more prone to lose track due to noise.

Precomputing. Computationally, the most expensive part of the algorithm was interpolating the values of $(T^l_\Delta p^l_t)$ on the shifted along the trajectory (and therefore no longer uniform) grid back to the original uniform grid. To speed up the computations, off-line part of the algorithm computes a (very sparse) matrix A such that multiplying the vector of values on the non-uniform grid by A produces the vector of values on the uniform grid.

Two interpolation algorithms were implemented: nearest neighbor interpolation (then matrix A has only zero and one as elements, and each row has at most one non-zero), and linear interpolation using known values on a triangle surrounding a point on the new grid. Linear interpolation proved to be much more stable and precise in simulations.

7.3.2. ***Simulation results.*** The simulation producing results included below used two targets and the noise with standard deviation $\sigma = 1.4$ which corresponds to SNR $= -2.9$dB.

- In Figure 17 ($t = 1s$), the target is present but not yet detected.
- In Figure 18 ($t = 6s$), a single sharp peak has formed, and it follows the target.
- In Figure 19 ($t = 11s$), the second target just appears at the horizon, algorithm doesn't track it yet.
- In Figure 20 ($t = 16s$), both targets were detected and are being tracked.

Also we refer to enclosed files (multiframe GIF file `a6.gif` or MS video file `a6.avi`) to see the incoming observation frames (Navy IRSS sensor – the LAPTEX field test) with very dim synthetic target inserted. The second pair of files (multiframe GIF file `a5.gif` or MS video file `a5.avi`) shows the locations of two targets being tracked and the filtering density which develops prominent peaks that follow both targets.

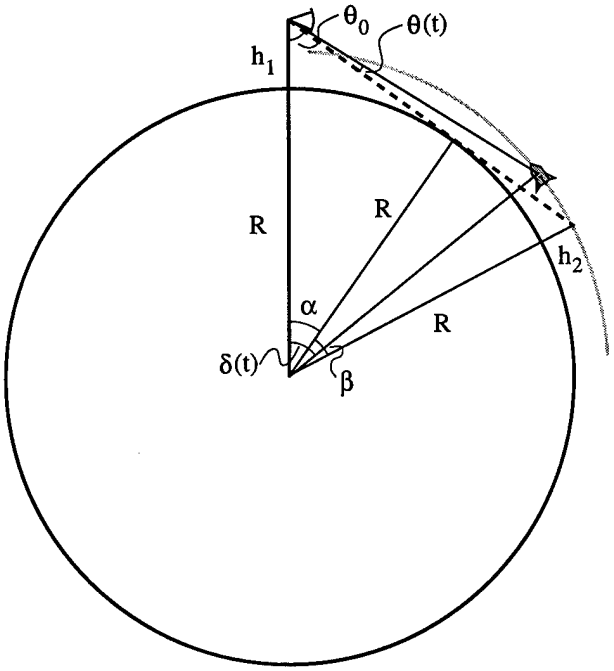


Figure 15. Surface skimming missile trajectory derivation

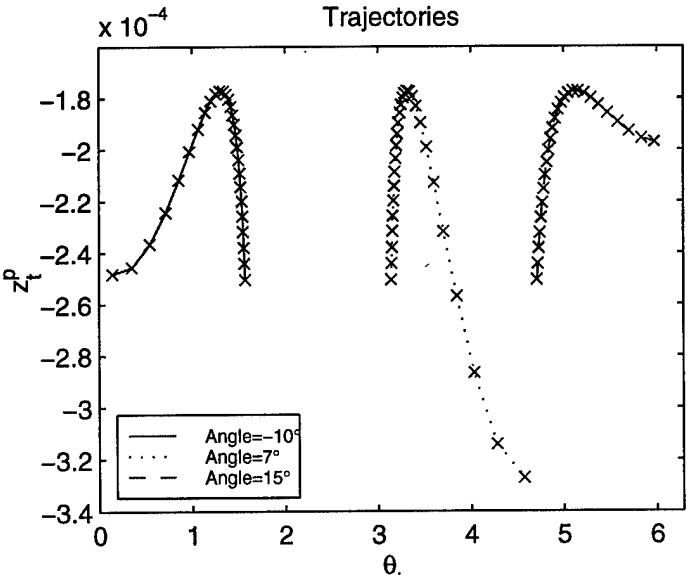


Figure 16. Surface skimming missile trajectories

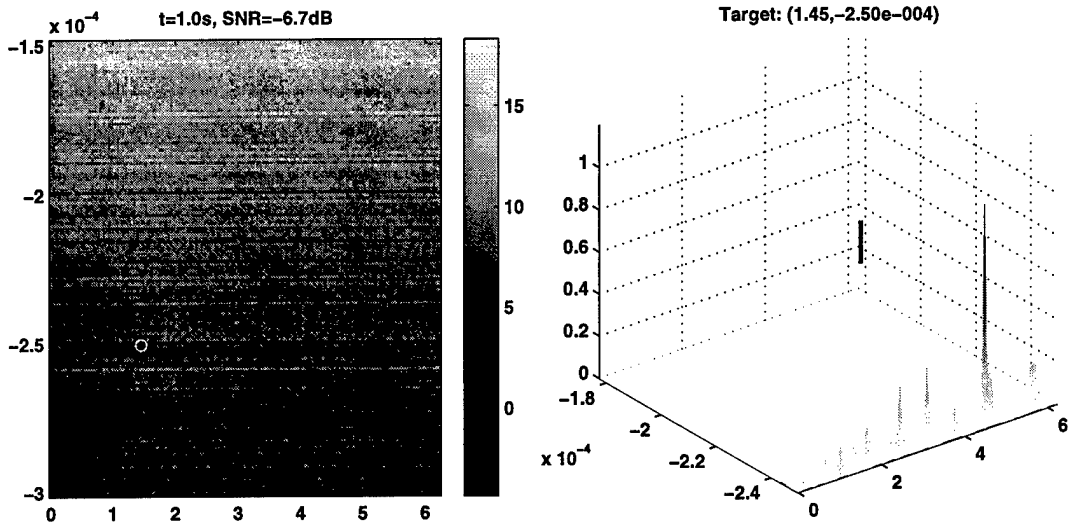


Figure 17. Surface skimming missile tracking, $t = 1s$

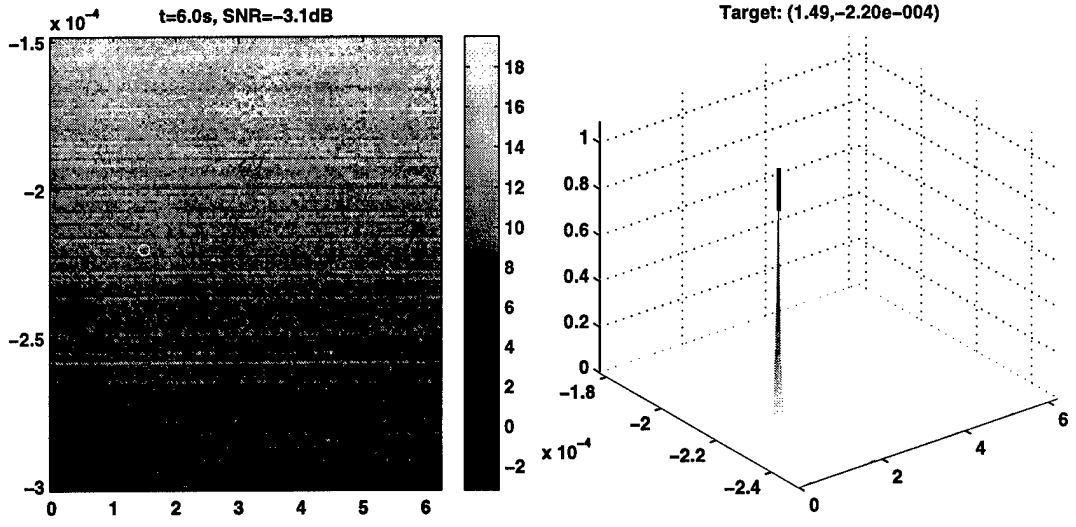


Figure 18. Surface skimming missile tracking, $t = 6s$

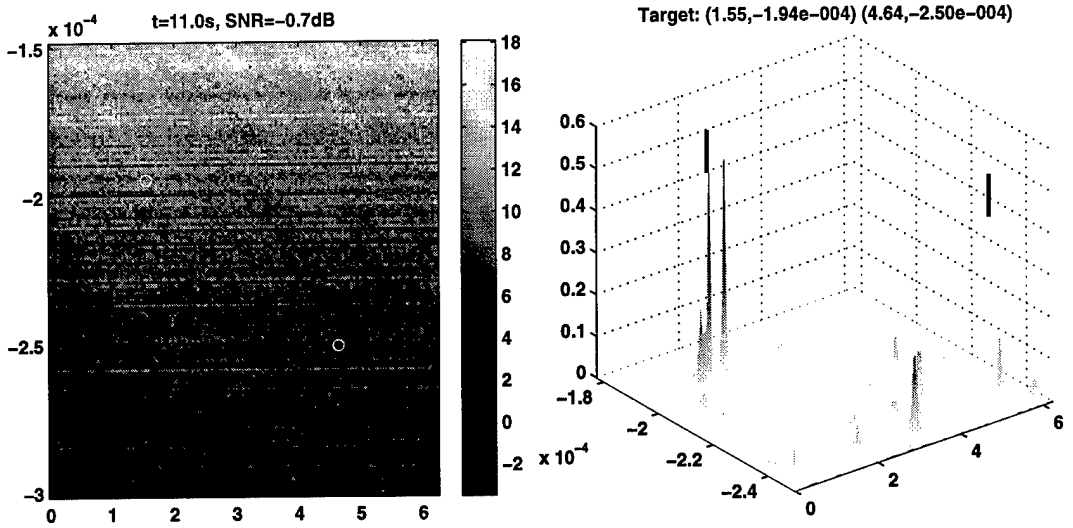


Figure 19. Surface skimming missile tracking, $t = 11s$

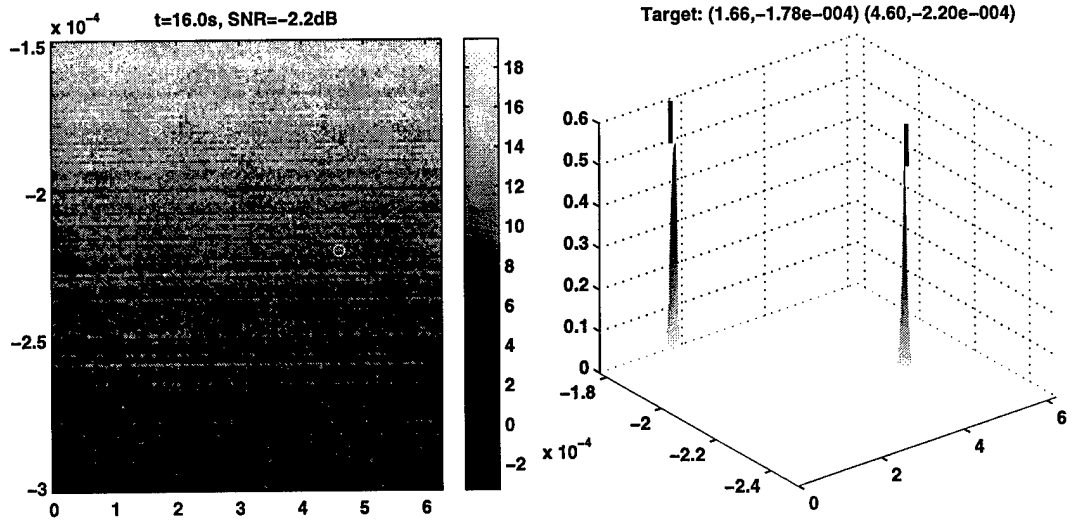


Figure 20. Surface skimming missile tracking, $t = 16s$

7.4. **Appearance/Disappearance Detection of a Skimming Missile.** In this section we present the results of application of the detection algorithms developed in Section 6. The model is the same as in Section 7.3. The results of TbD (the estimates (7.3) and (7.4) of the signal location) are used in the LLR estimate (6.43). In the fully sequential CUSUM algorithm, the adaptive CUSUM statistic U_n^* is computed recursively according to (6.44) and then is compared to the threshold a (see (6.45)). In the sequential (adaptive) quasi-Bayesian algorithm the estimate of the logarithm of average likelihood ratio $r_n^* = \log R_n^*$ is computed according to the recursive formula (6.47). It is then compared to the threshold $\log B$ (see (6.46)).

Figure 21, Figure 22, and Figure 23 illustrate the behavior of the adaptive CUSUM statistic for different S(N+C)R (SNR = −0.25dB, SNR = −3.62dB, and SNR = −6.6dB, respectively). Theoretically $0 \le r_n^* - U_n^* \le \log n$. We have not observed any substantial difference between these statistics in our experiments – the difference was always negligible. So we omit the plots of the statistic r_n^* and show only the behavior of the adaptive CUSUM. It turns out that the mean estimate (7.3) provides better results than the maximum posterior estimate (7.4) and we plot only the CUSUM statistic that uses the former one. It is seen that most of the time the statistic U_n^* is close to zero or negative when the target is absent. Sometimes, however, peaks arise. These peaks may be identified with signal presence. But they are typically short and may be easily distinguished from the peaks due to target. One possible method to discriminate between false alarms and true target is to make the decision on target presence if there are several subsequent exceedances (say 3-4) of the threshold. Otherwise the decision on target absence is made (i.e. a single exceedance is identified with false alarm). When the target appears the statistic rapidly increases while when it disappears, U_n^* decreases. In our experiments we observed visible difference in behavior of U_n^* when the signal appears compared to the case of its absence up to the SNR −6.6dB. In the pictures the first target appears at time $n = 1$ and disappears at $n = 28$. The second target appears at time $n = 39$ and never disappears.

Figure 24 shows the results of detection of track appearance and disappearance by Algorithm 3 for SNR = −0.25dB and SNR = −6.6dB. The detection of tracks occurs when the adaptive CUSUM U_n^* exceeds the threshold a (the upper one) and track disappearance is detected when the statistic $\Delta_n^* = L_{n-1}^* - U_n^*$ exceeds the threshold b (the lower one). When the decision on target disappearance is made, the statistic U_n^* is renewed form zero, i.e. the detection algorithm is prepared to detect another target. It is seen that the algorithm is able to detect even very low SNR targets. The threshold a was chosen such that the false alarm rate 1 false alarm per minute (i.e., per 60 frames) was guaranteed. The algorithm never raised more than 1 false alarm and always detected targets with this choice in the analyzed images. To be precise, there are two targets in the pictures: the first target appears at time $n = 1$ and disappears at $n = 28$ while the second one appears at time $n = 39$ and does not disappear. The algorithm detects the first target with the delay about 20 seconds (20 frames) for SNR = −6.6dB and $4 - 6$ frames for SNR = −0.25dB. Then the fact of target's disappearance is detected with very small delay. Finally the second target is detected with the delay about $5 - 6$ frames for SNR = −6.6dB and $2 - 3$ frames for SNR = −0.25dB.

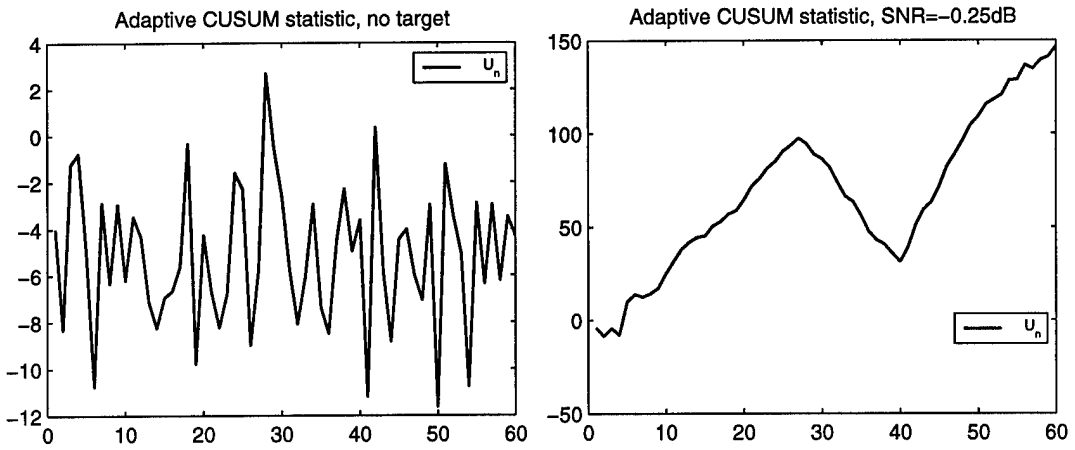


Figure 21. Plots of the adaptive CUSUM. Left – no target. Right – 1st target appears at $n = 1$, disappears at $n = 28$; 2nd appears at $n = 39$, doesn't disappear

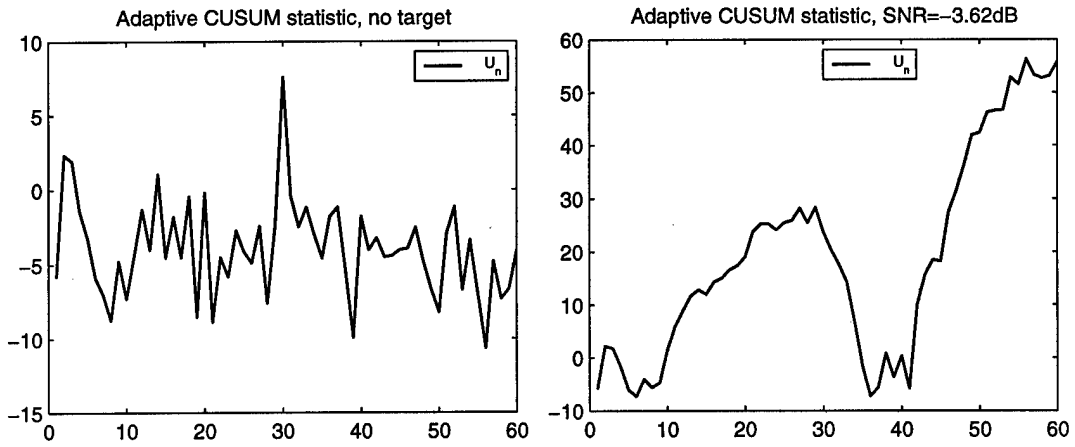


Figure 22. Plots of the adaptive CUSUM. Left – no target. Right – 1st target appears at $n = 1$, disappears at $n = 28$; 2nd appears at $n = 39$, doesn't disappear

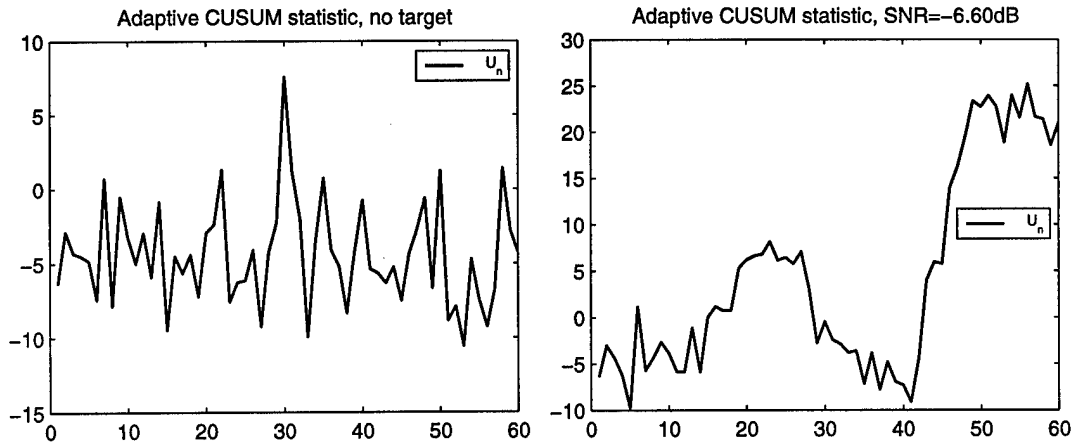


Figure 23. Plots of the adaptive CUSUM. Left – no target. Right – 1st target appears at $n = 1$, disappears at $n = 28$; 2nd appears at $n = 39$, doesn't disappear

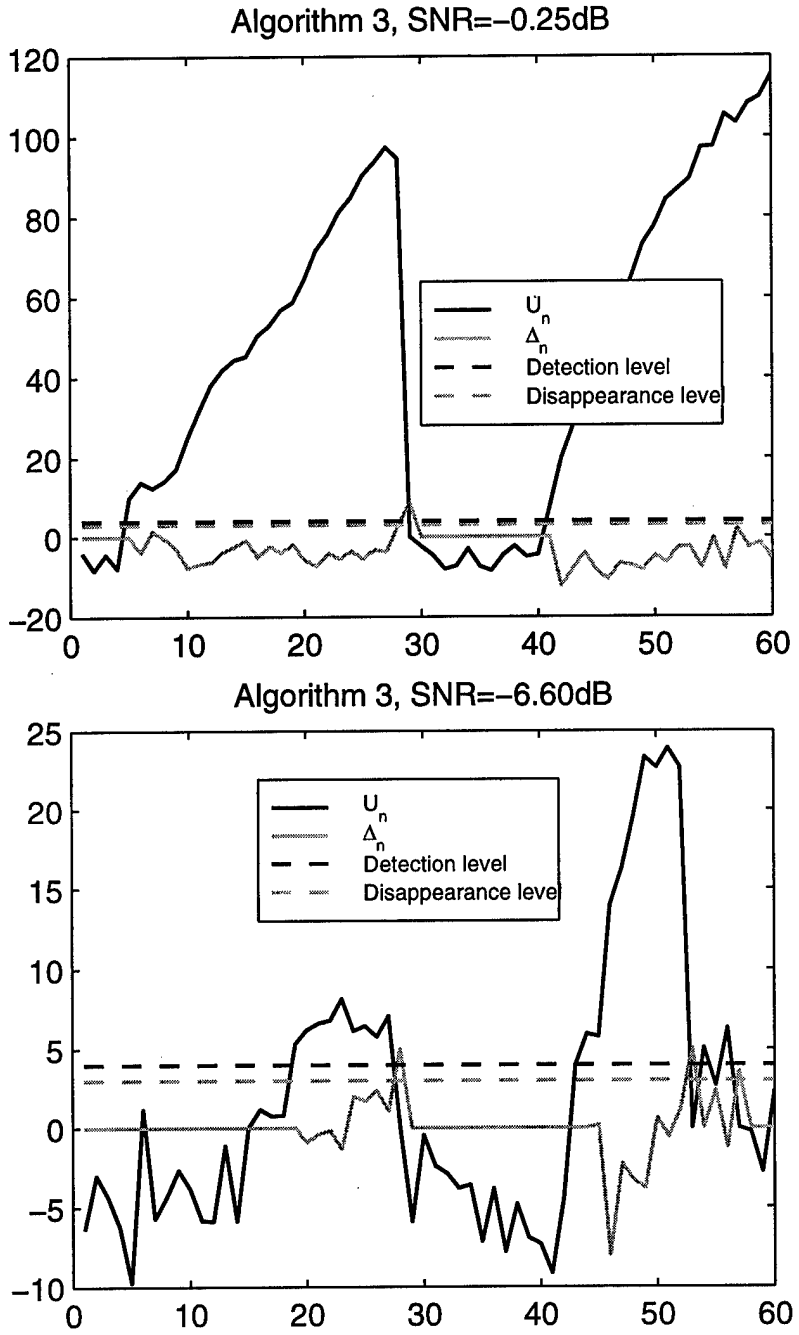


Figure 24. Detection of track appearance and disappearance by Algorithm 3. First target appears at $n = 1$ and disappears at $n = 28$; second target appears at $n = 39$ and doesn't disappear

8. Conclusion. Future Research

We believe that real time optimal nonlinear filtering is an emerging information technology of fundamental importance. The ONF is an advance over conventional EKF techniques and multi-dimensional (spatio-temporal) matched filters potentially as profound as that of the EKF over $\alpha-\beta-\gamma$ filters. Nonlinear filtering techniques offer the promise of greatly enhanced performance in a number of DoD related applications: airborne, surface, and subsurface.

We have developed an advanced ONF-based spatio-temporal tracking filter and adaptive track appearance/disapperance detection algorithms that may be fully incorporated into a complete end-to-end IRST signal/track processing suite. The developed technology will be of portable nature and can be incorporated later in various multisensor ATR systems utilizing EO/IR, SAR, SAS, LIDAR, etc.

In the future we plan to continue this work in the following key directions.

1. Further develop and tune spatio-temporal nonlinear semiparametric algorithms for clutter removal and jitter compensation.
2. Develop banks of ONF filters for TbD of multiple targets.
3. Improve and tune track detection and identification algorithms.
4. Test and validate the developed signal/data/track processing system for real IR data (jointly with Space & Naval Warfare Systems Center, San Diego, CA).
5. Incorporate the developed system into a complete end-to-end IRST signal/track processing suite (jointly with Space & Naval Warfare Systems Center, San Diego, CA).

9. Acknowledgement

We are grateful to Dr. John Barnett of Space & Naval Warfare Systems Center, San Diego, CA for very useful discussions and for providing the real IR data used in the analysis.

References

[1] A. Aridgides, M. Fernandez, and D. Randolph, "Adaptive three-dimensional spatio-temporal IR clutter suppression filtering techniques," *SPIE Proc.*, Vol. 1305, 1990.

[2] J. Arnold and H. Pasternack, "Detection and tracking of low-observable targets through dynamic programming," *SPIE Proc.*, Vol. 1305, 1990.

[3] P. Athanas and A. Abbott, "Real-time image processing on a custom computing platform," *IEEE Computer*, pp. 16-24, 1995.

[4] J.T. Barnett, B.D. Billard, and C. Lee, " Nonlinear morphological processors for point-target detection versus an adaptive linear spatial filter: a performance comparison," *Proceedings of SPIE Symposia on Aerospace Sensing*, Vol. 1954, pp. 12-24, Orlando, FL, 1993.

[5] J.T. Barnett, "Statistical analysis of median subtraction filtering with application to point target detection in infrared backgrounds," *Proc. SPIE Symposium on Infrared Signal Processing*, Vol. 1050, pp. 10-18, Los Angeles, CA, 1989.

[6] C. Barlow and S. Blackman, "New Bayesian track-before-detect design and performance study," *SPIE Proc.*, Vol. 3373, 1998.

[7] Y. Barniv, "Dynamic programming solution for detecting dim moving targets," *IEEE Trans. AES*, Vol. 21, 1985.

[8] Y. Bar-Shalom and T.E. Fortmann, *Tracking and Data Association* Orlando: Academic Press, 1988.

[9] S. Blackman, R. Dempster, and T. Broida, "Multiple hypothesis track confirmation for infrared surveillance systems," *IEEE Trans. AES*, Vol. 29, pp. 810-823, 1993.

[10] S.S. Blackman, "Low SNR detection and tracking," Technical Report, Hughes Aircraft Co., 1994.

[11] S.S. Blackman, *Multiple Target Tracking with Radar Applications.* Artech House, 1986.

[12] H. Blom, "An efficient filter for abruptly changing systems," *Proc. 23rd IEEE Conf. Decision Contr.*, pp. 656-658, 1984.

[13] H.A.P. Blom, et al., "Design of a multisensor tracking system for advanced air traffic control," In: *Multitarget-Multisensor Tracking*, Vol. 2 (ed. Y. Bar-Shalom), pp. 31-63, 1992.

[14] H. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Autom. Contr.*, Vol. 33, pp. 780-783, 1988.

[15] G.C. Demos, R.A. Ribas, T.J. Broida, and S.S. Blackman, "Applications of MHT to dim moving targets," *Proc. SPIE: Signal and Data Processing of Small Targets*, Vol. 1305, pp. 297-309, 1990.

[16] Y. Chan, J. Plant, and J. Bottomley, "A Kalman tracker with a simple input estimator," *IEEE Trans. AES*, Vol. 18, pp. 235-241, 1982.

[17] L.J. Cutrona, "Synthetic Aperture Radar," In: *Radar Handbook* (Ed. M. Skolnik). New York: MacGraw-Hill, 1990.

[18] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet Shrinkage: Asymptopia ?" (with discussion) *J. Roy. Stat. Soc. B*, Vol. 57, no. 1, pp. 301-369, 1995.

[19] D.L. Donoho, "Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery," *Probab. Theory Rel. Fields*, Vol. 99, pp. 145-170, 1994.

[20] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2. New York: John Wiley & Sons, 1966.

[21] M. Fernandez, A. Aridgides, and D. Bray, "Detecting and tracking low-observable targets using IR," *SPIE Proc.*, Vol. 1305, 1990.

[22] M.A. Girshik and H. Rubin, "A Bayes approach to a quality control model," *Ann. Math. Statist.*, Vol. 23, pp. 114-125, 1952.

[23] W. Härdle, *Applied Nonparametric Regression.* Cambridge: Cambridge University Press, 1990.

[24] A.H. Jazwinski, *Stochastic Processes and Filtering Theory.* New York: Academic Press, 1970.

[25] S. Kligys and B.L. Rozovskii, "Optimal nonlinear filtering with distributed observation," CAMS Preprint, USC, 1997.

[26] S.L. Leonov, "On the solution of an optimal recovery problem and its applications in nonparametric regression," *Mathematical Methods of Statistics*, Vol. 6, no. 4, pp. 476-490, 1997.

[27] O.V. Lepski, E. Mammen, and V.G. Spokoiny, "Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors," *Ann. Stat.*, Vol. 25, pp. 929-947, 1997.

[28] R. Lindgren and L. Taylor, "Bayesian field tracking," *SPIE Proc.*, Vol. 1954, 1993.

[29] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, Vol. 42, pp. 1897-1908, 1971.

[30] S.V. Lototsky, R. Mikulevicius, and B.L. Rozovskii, "Nonlinear filtering revisited: a spectral approach," *SIAM J. Control Optim.*, Vol. 35, no. 2, 1997, pp. 435-461.

[31] S.V. Lototsky and B.L. Rozovskii, "Recursive nonlinear filter for a continuous-discrete time model: separation of parameters and observations," *IEEE Trans. Autom. Control*, Vol. 43, no. 8, pp. 1154-1158, 1998.

[32] S.V. Lototsky, C. Rao, and B.L. Rozovskii, "Fast nonlinear filter for continuous-discrete time multiplication models," *Proc. 35th Conf. Decision and Control*, Kobe, Japan, 1996, Omnipress, Madison, Wisconsin, Vol. 4, pp. 4060-4064.

[33] S.V. Lototsky, C. Rao, and B.L. Rozovskii, "Method of optimal stochastic filtering for tracking objects with possibly nonlinear dynamics," *U.S. Patent* (pending).

[34] N. Mohanty, "Computer tracking of moving point targets in space," *IEEE Trans. PAMI*, Vol. 3, 1981.

[35] E.S. Page, "A test for a change in a parameter occurring at an unknown point," *Biometrika*, Vol. 42, no. 4, pp. 523-527, 1955.

[36] N. Platt, S. Hammel, J. Trahan, and H. Rivera, "Mirages in the marine boundary layer – comparison of experiment with model," *Proc. IRIS Passive Sensors*, Monterey, CA, Vol. 2, pp. 195-210, 1996.

[37] N. Platt and S.M. Hammel, "Pattern formation in driven coupled map lattices," *Physica A*, Vol. 239, pp. 296-303, 1997.

[38] M. Pollak, "Optimal detection of a change in distribution," *Ann. Statist.*, Vol. 13, pp. 206-227, 1986.

[39] I.S. Reed, R.M. Gagliardi, and H.M. Shao, " Application of three-dimensional filtering to moving target detection," *IEEE Trans. AES*, Vol. 19, 1983, pp. 898-904.

[40] I.S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 38, pp. 1760-1770, 1990.

[41] V.G. Repin, "Detecting a signal with unknown moments of appearance and disappearance," *Problems Inform. Transmission*, Vol. 27, no. 1, pp. 61-72. 1991.

[42] S.W. Roberts, "A comparison of some control chart procedures," *Technometrics*, Vol. 8, no. 3, pp. 411-430, 1966.

[43] B.L. Rozovskii, *Stochastic Evolution Systems: Linear Theory and Application to Nonlinear Filtering*, MIA, Kluwer Academic Publishers, 1990.

[44] J. Sanders, "A method for determining filter spacing in assumed velocity filter banks," *IEEE Trans. AES*, Vol. 29, 1993.

[45] A.N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Probab. Applications*, Vol. 8, pp. 22-46, 1963.

[46] W.W. Shrader and V. Gregeres-Hansen, "MTI Radar," In: *Radar Handbook* (Ed. M. Skolnik). New York: MacGraw-Hill, 1990.

[47] D. Siegmund, *Sequential Analysis.* New York: Springer-Verlag, 1985.

[48] P.F. Singer and D.M. Sasaki, "The heavy tailed distribution of a common CFAR detector," *SPIE Proc.*, Vol. 2561, pp. 124-140, 1995.

[49] P. Singer, "Performance analysis of a velocity filter bank," *SPIE Proc.*, Vol. 3163, 1997.

[50] F.M. Staudaher, "Airborne MTI," In: *Radar Handbook* (Ed. M. Skolnik). New York: MacGraw-Hill, 1990.

[51] A. Stocker and P. Jensen, "Algorithms and architectures for implementing large velocity filter banks," *SPIE Proc.*, Vol. 1481, 1991.

[52] A.G. Tartakovsky and S. Blackman, "Asymptotically optimal sequential tests and application to target detection and discrimination," *30th Annual Conference on Information Sciences and Systems*, Princeton, NJ, 1996.

[53] A.G. Tartakovsky, *Sequential Methods in the Theory of Information Systems.* Moscow: Radio i Svyaz', 1991.

[54] A.G. Tartakovskii, "Efficiency of the generalized Neyman-Pearson test for detecting changes in a multichannel system," *Problems of Information Transmission*, Vol. 28, pp. 341-350, 1992.

[55] A.G. Tartakovskii, "Asymptotic properties of CUSUM and Shiryaev's procedures for detecting a change in a nonhomogeneous Gaussian process," *Mathematical Methods of Statistics*, Vol. 4, no. 4, pp. 389-404, 1995.

[56] A.G. Tartakovsky, "Minimal time detection algorithms and applications to flight systems," School of Engineering and Applied Science, Flight Systems Research Center, Technical Report# 2-FSRC-93, 1993.

[57] A.G. Tartakovsky, "Detection of signals with random moments of appearance and disappearance," *Problems of Information Transmission*, Vol. 24, pp. 115-124, 1988.

[58] S. Tonissen and R. Evans, "Performance of dynamic programming techniques for track-before-detect," *IEEE Trans. AES*, Vol. 32, 1996.

[59] P. Wei, J. Zeidler, and W. Ku, "Analysis of multiframe target detection using pixel statistics," *IEEE Trans. AES*, Vol. 31, 1995.

[60] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis.* Philadelphia: SIAM, 1982.