A Particle Filtering Approach to Abnormality Detection in Nonlinear Systems and its Application to Abnormal Activity Detection *

Namrata Vaswani, Rama Chellappa

Center for Automation Research

Department of Electrical and Computer Engineering

University of Maryland, College Park

MD 20742, USA

{namrata,rama}@cfar.umd.edu

Abstract

We study abnormality detection in partially observed nonlinear dynamic systems tracked using particle filters. An 'abnormality' is defined as a change in the system model, which could be drastic or gradual, with the parameters of the changed system unknown. If the change is drastic the particle filter will lose track rapidly and the increase in tracking error can be used to detect the change. In this paper we propose a new statistic for detecting 'slow' changes or abnormalities which do not cause the particle filter to lose track for a long time. In a previous work, we have proposed a partially observed nonlinear dynamical system for modeling the configuration dynamics of a group of interacting point objects and formulated abnormal activity detection as a change detection problem. We show here results for abnormal activity detection comparing our proposed change detection strategy with others used in literature.

* Partially supported by the DARPA/ONR Grant N00014-02-1-0809

1 Introduction

In many practical problems arising in quality control or surveillance problems like abnormal activity detection, the underlying system in its normal state can be modeled as a parametric stochastic model (which may be linear or nonlinear). In most cases, the observations are noisy (making the system partially observed) and the transformation between the observation and the state may again be linear or nonlinear. Such a system, in the most general case, forms a Partially Observed Non-Linear Dynamical (PONLD) system which can be tracked using a particle filter [1]. An 'abnormality' constitutes a change, which could be drastic or gradual, in the system model, with the parameters of the system model after the change unknown. If the change in the system is 'drastic' (or abrupt) the particle filter will lose track rapidly and the increase in tracking error can be used to detect the change [2]. In this paper we propose a new strategy for detecting *'slow' (gradual) changes* or abnormalities in partially observed *nonlinear systems* which do not cause the particle filter to lose track for very long and whose *parameters are unknown*, and show its application to abnormal activity detection. In a previous work [3], we have modeled the changing configuration of a group of moving point objects (an "activity") as a moving deformable shape in $\Re^n$. We have proposed a PONLD system to model an activity with the learnt shape deformation and motion models forming the system model and the measurement noise in the object locations (configuration space) forming the observation model. The mapping from state (shape+motion) space to observation (configuration) space is nonlinear. In this paper, we apply our proposed change detection strategy for abnormal activity (suspicious behavior) detection in this framework and compare its performance, for both slow and drastic abnormalities, with other statistics used in literature.

1.1 Previous Work

Online detection of changes for partially observed *linear* dynamical systems has been studied extensively. For *known changed system parameters* the CUSUM [4] algorithm can be used directly. The CUSUM algorithm uses as change detection statistic, the maximum (taken over all previous time instants) of the likelihood ratio assuming change occurred at time j, i.e. $CUS_t \triangleq \max_{1 \leq j \leq t} \frac{p_{\theta_1}(y_j, y_{j+1} \dots y_t)}{p_{\theta_0}(y_j, y_{j+1} \dots y_t)}$. For *unknown changed system parameters*, the Generalized LRT can be used whose solution for *linear* systems in well known [4]. When a nonlinear system experiences a change, linearization techniques like

Extended Kalman Filtering [5] and change detection methods for linear systems are the main tools for change detection [4]. Linearization techniques are computationally efficient but are not always applicable (require a good initial guess at each time step and hence not robust to noise spikes).

[6] is an attempt to use a Particle Filtering (PF) approach for *abrupt* change detection in Partially Observed *Non-Linear* Dynamical (PONLD) systems without linearization. It uses *knowledge of the parameters of the changes system* and defines a modification of the CUSUM change detection statistic that can be evaluated recursively. It runs a sequence (for different j) of PF's to evaluate the probability of the observations given that the change occurred at time j and compares each with the observations' probability evaluated using a PF which assumes that no change occurred. Another statistic commonly used for *abrupt* change detection in partially observed systems and which *does not require knowledge of changed system parameters* is tracking error [2]. Tracking error is the distance (usually Euclidean distance) between the current observation and its prediction based on past observations. Another class of approaches (eg. see [7, 8]) used extensively with particle filters uses a *discrete state variable* to denote the mode that the system is operating in. This is typically used when the system can operate in multiple modes each associated with a different system and/or observation model. The mode variable's transition between states is governed by which mode maximizes the likelihood of the observations. But this approach also assumes *known* change parameters.

Particle filtering (PF) [9] also known as the sequential Monte Carlo method [10] or condensation algorithm [11] provides a finite dimensional approximate solution to the conditional density of the state given past observations. In this paper, we use a PF (described in section 3) to track the PONLD system. The PF provides at each time instant t, an n-sample monte carlo estimate of the poterior distribution of the state of the system, given observations upto time t. We propose to use as a change detection statistic, the expectation under this posterior state distribution of the log of the a-priori pdf of the state at time t and we call this statistic ”Expected Log Likelihood” or ELL. This is discussed in section 4. In section 5, we prove convergence of the Huber M-estimate [12] of this statistic, evaluated using the posterior distribution estimated by the PF, to its true value. Also, to be able to detect both slow and drastic changes, we propose to use a combination of ELL and tracking error. Finally in section 6 we show results for abnormal activity detection using ELL, tracking error, observation likelihood, $[-\log Pr(y_t|y_{0:t-1})]$, and the combined ELL-tracking error strategy and discuss future work in section 7.

2 Problem Statement and Assumptions

Given a partially observed nonlinear dynamic (PONLD) system [13] with an $\Re^{n_x}$ valued state process $X = \{X_t\}$ and an $\Re^{n_y}$ valued observation process $Y = \{Y_t\}$, the aim is to detect a change in the system model (model for $\{X_t\}$). The system (or state) process $\{X_t\}$ for the normal system is assumed to be a Markov process with state transition kernel $K_t(x_t, dx_{t+1})$ and the observation process is defined by $Y_t = h_t(X_t) + w_t$ where w_t is an i.i.d. noise process and h_t is in general a nonlinear function. The prior initial state distribution denoted by $\pi_0(dx)$, the observation likelihood denoted by $G_t(dy_t|x_t)$ and the state transition kernel $K_t(x_t, dx_{t+1})$ are known. We assume in this work that the prior $\pi_0(dx)$ is absolutely continuous (w.r.t. to the Lebesgue measure) i.e. it admits a density (pdf) $p_0(x)$, the transition kernel K_t is Feller [13] and also absolutely continuous for all values of x_t and for all t. Also the observation likelihood is absolutely continuous i.e. it admits a density $g_t(y_t|x_t)$ which is a continuous and everywhere positive function.

Abnormality detection in a *PONLD system* is posed as a change detection problem with *parameters of the changed system unknown*. If the abnormality is a drastic one, it would cause the PF to lose track (tracking error will increase) and hence will get detected. But in a lot of cases, the abnormality is a *slow change* that does not cause the PF to lose track for a long time (there will be a large detection delay if tracking error is used) and this is the problem that we address here. To be specific, the problem is to detect a gradual change in the system model i.e. detect with minimum delay, the time instant t_{change} when the state starts following a different system model with transition kernel $\tilde{K}_t(x_t, dx_{t+1})$ which is unknown.

3 Particle filtering (PF)

The particle filter [13] is a recursive algorithm which produces at each time t, a cloud of n particles, $\{x_t^{(i)}\}_{i=1}^n$, whose empirical measure closely "follows" $\pi_t(dx_t|y_{0:t})$ the posterior distribution of the state given past observations (denoted by $\pi_{t|t}(dx_t)$ in the rest of the paper). It starts with sampling n times from the initial state distribution $\pi_0(dx)$ to approximate it by $\pi_0^n(dx) = \frac{1}{n}\sum_{i=1}^n \delta_{x_0^{(i)}}(dx)$ and then implements the Bayes' recursion [13] at each time step. Now assuming that the distribution of X_{t-1} given observations upto time $t-1$ has been approximated as $\pi_{t-1|t-1}^n(dx) = \frac{1}{n}\sum_{i=1}^n \delta_{x_{t-1}^{(i)}}(dx)$, the prediction step samples the new state $\bar{x}_t^{(i)}$ from the distribution $K_{t-1}(x_{t-1}^{(i)}, .)$. Thus the empirical distribution of this new cloud of particles, $\pi_{t|t-1}^n(dx) = \frac{1}{n}\sum_{i=1}^n \delta_{\bar{x}_t^{(i)}}(dx)$ is an approximation to the probability distribution of X_t given observations upto time $t-1$. For each particle, its weight is proportional to the likelihood of the observation given that particle, i.e. $w_t^{(i)} = \frac{n g_t(y_t|\bar{x}_t^{(i)})}{\sum_{i=1}^n g_t(y_t|\bar{x}_t^{(i)})}$. $\bar{\pi}_{t|t}^n(dx) = \frac{1}{n}\sum_{i=1}^n w_t^{(i)}\delta_{\bar{x}_t^{(i)}}(dx)$ is then an estimate of the probability distribution of the state at time t given observations uptil time t. We resample n times with replacement from $\bar{\pi}_{t|t}^n(dx)$ to obtain the empirical estimate $\pi_{t|t}^n(dx) = \frac{1}{n}\sum_{i=1}^n \delta_{x_t^{(i)}}(dx)$. Note that both $\bar{\pi}_{t|t}^n$ and $\pi_{t|t}^n$ approximate $\pi_{t|t}$ but the resampling step is used because it increases the sampling efficiency by eliminating samples with very low weights.

4 Change/ Abnormality Detection

First let us define some notation. The integral of a function c w.r.t. a measure ρ is denoted by (ρ, c) with

$$(\rho, c) = \int c(x)\rho(dx) \tag{1}$$

Given the prior initial state distribution π_0 and the transition kernel K_t, the prior state distribution at any time t is $(((\pi_0, K_0), K_1), ...K_t)$. Since the transition kernel is absolutely continuous, this state distribution admits a pdf $p_t(x)$. In a lot of cases (for example if the system model is linear Gaussian with Gaussian initial state pdf) it is possible to define the pdf $p_t(x)$ in closed form. Now, if the system were fully observed, one could evaluate $x_t = {h_t}^{-1}(y_t)$ from the observation y_t and then $[-\log p_t(x_t)]$ (negative log-likelihood of the state coming from a normal system) could be used as a change detection statistic; the negative log-likelihood of the state of the changed or abnormal system would be smaller than that of the normal system.

But for a partially observed system we can only approximate (using a PF) the posterior distribution of the state given past observations, $\pi_{t|t}(dx_t)$ by $\pi_{t|t}^n(dx) \triangleq \frac{1}{n}\sum_{i=1}^n \delta_{x_t^{(i)}}(dx)$ which has Radon Nikodym density $\pi_{t|t}^n(x) = \frac{1}{n}\sum_{i=1}^n \delta_{x_t^{(i)}}(x)$. The negative log-likelihood of state coming from a normal system can now be replaced by its expectation under the posterior distribution of the state. Thus for the partially observed system tracked using a PF, we propose to use as the change detection statistic, the expected log-likelihood (ELL), $E_{\pi_{t|t}}[-\log p_t(X)] = (\pi_{t|t}, [-\log p_t(X)])$, approximated by $(\pi_{t|t}^n, [-\log p_t(X)])$.

It is interesting to note that ELL is also the **Kerridge inaccuracy** [14] between $[\frac{d\pi_{t|t}}{dx}]$ (the pdf corresponding to $\pi_{t|t}$) and p_t, i.e. $E_{\pi_{t|t}}[-\log p_t(X)] = K([\frac{d\pi_{t|t}}{dx}] : p_t)$. This is approximated using the

PF as $K(\pi^n_{t|t} : p_t)$. The inaccuracy between the posterior state distribution corresponding to normal observations and the prior will be smaller than that between the posterior corresponding to abnormal observations and the prior. Kerridge inaccuracy can also be interpreted as the average length of code designed for $p_t(x_t)$ rather than $\pi_{t|t}(dx_t)$ [15].

The advantage of this statistic over others used in literature (for PONLD systems tracked using PF's) like tracking error [2] or log-likelihood of current observation given past observations [6] is that it can detect slow changes much better. Consider as an example to motivate our approach, a PONLD system with a linear Gaussian system model i.e. $x_t = Ax_{t-1} + n_t$. n_t is i.i.d. zero mean Gaussian noise, $p_0(x_0)$ is a zero mean Gaussian, so that $p_t(x)$ is also a zero mean Gaussian density. Let us assume that the abnormality causes the system model to change to $\tilde{x}_t = A\tilde{x}_{t-1} + n_t + b_t$, $\forall t \geq t_{change}$, where $b_t = b$ is a small constant bias added to the state or part of the state vector. If the bias added along a direction is small or comparable to the noise variance along that direction, it will get tracked correctly by the PF and hence tracking error will not show a significant change. But a systematically increasing bias along certain directions will cause the mean of the posterior distribution of the state to increase to a significantly nonzero value within a few time steps, thus causing the Kerridge inaccuracy between the posterior and the normal system pdf to increase. For such a problem, ELL (or Kerridge inaccuracy) will detect the abnormality much faster than tracking error.

Now, if the abnormality/ change is a drastic one, it will cause the PF to lose track quite rapidly (the number of particles, n, is a finite number, and so there will be very few particles around the expected value of the abnormal state), thus causing the tracking error to show a sudden increase which can be used for detecting an abrupt change. Due to the same reason, the estimate of the posterior distribution in this case is no longer reliable and hence the ELL evaluated with it will also be inaccurate. For this case tracking error is a more reliable change detection statistic. Hence a robust change detection strategy which can detect both slow and drastic changes with minimum delay would be to use both tracking error and ELL to detect a change; if either one exceeds its respective threshold then a change is declared.

5 Convergence of Approximation

Let $(\Omega, \mathcal{F}, P)$ be a probability space, then the sequence $\pi^n_{t|t}$ is a sequence of random measures, $\pi^n_{t|t} : \Omega \rightarrow \mathcal{P}(\Re^{n_x})$ and $\pi_{t|t} \in \mathcal{P}(\Re^{n_x})$ is a deterministic probability measure. $\mathcal{P}(\Re^{n_x})$ denotes the set of probability measures over the Borel sigma algebra on $\Re^{n_x}$.

Theorem 1 *[13] : The sequence $\pi^n_{t|t}$ of (empirical) posterior state distributions estimated using a PF (that satisfies certain smoothness assumptions discussed in [13]) converges weakly to the true posterior $\pi_{t|t}$, P-a.s., i.e.*

$$\lim_{n\to\infty}(\pi^n_{t|t}, c) = (\pi_{t|t}, c), \text{ P-a.s. } \forall c \in \mathcal{C}_b(\Re^{n_x}) \tag{2}$$

where $\mathcal{C}_b(\Re^{n_x})$ denotes the set of continuous bounded functions on $\Re^{n_x}$.

First note that the PF used in this work satisfies the assumptions of [13]: As discussed earlier in section 2, the observation likelihood, is a continuous and everywhere positive funciton and the transition kernel is Feller. The multinomial resampling step (described in section 3) produces n unbiased samples from $\bar{\pi}_{t|t}$ with variance of the weights estimated from the samples bounded by $1/n$ (so it goes to zero as $n \to \infty$).

Now, the negative log likelihood of state $[-\log p_t(x)] \triangleq f(x)$ is an unbounded function while the theorem stated above works for bounded continuous functions. But since it is a non-negative function, we can use the standard mathematical analysis trick of approximating it by a sequence of increasing bounded functions which will converge to f, i.e. define

$$f_m(x) = \min(f(x), m) = \min([-\log p_t(x)], m) \tag{3}$$

Then we have $\lim_{m\to\infty} f_m(x) = f(x), \forall x$. Now, f_m is a continuous bounded function and so we can use theorem 1 with $c = f_m$ to prove convergence of $\mathcal{E}_{m,n} \triangleq (\pi^n_{t|t}, f_m)$ to $\mathcal{E}_m \triangleq (\pi_{t|t}, f_m)$, P-a.s., as $n \to \infty$, i.e.

$$\lim_{n\to\infty}(\pi^n_{t|t}, f_m) = (\pi_{t|t}, f_m) \triangleq \mathcal{E}_m, \text{ P-a.s, } \forall m. \tag{4}$$

Since f_m is a sequence of non-negative increasing functions that converge pointwise to f, we can use Monotone Convergence Theorem ([16], page 87) to prove convergence of $\mathcal{E}_m$ to $\mathcal{E} \triangleq (\pi_{t|t}, f)$, i.e.

$$\lim_{m\to\infty} \mathcal{E}_m = \lim_{m\to\infty}(\pi_{t|t}, f_m) = (\pi_{t|t}, f) \triangleq \mathcal{E} \tag{5}$$

Combining equations (4) and (5), we get

Corollary 1

$$\lim_{m\to\infty}(\lim_{n\to\infty}(\pi^n_{t|t}, f_m)) = (\pi_{t|t}, f), \text{ P-a.s.} \tag{6}$$

or that the error between the posterior expectation of the bounded approximation of negative log-likelihood of state, evaluated using an n-particle PF, and its true value can be made as small as one wants by choosing the bound to be large enough and given the bound, choosing n to be large enough.

Remark 1 *: Interpreting the above result in terms of* **Kerridge inaccuracy**, *we have*

$$\begin{aligned} \lim_{m\to\infty}(\lim_{n\to\infty} K(\pi^n_{t|t} : p^m_t)) &= K(\frac{d\pi_{t|t}}{dx} : p_t), \text{ P-a.s. where} \\ p^m_t(x) &\triangleq \max(p_t(x), e^{-m}), \text{ Note that } p^m_t \text{ is not a pdf since it does not intgrate to 1} \end{aligned} \tag{7}$$

Remark 2 *: $\mathcal{E}_{m,n}$ is evaluated using $\pi^n_{t|t}$, as follows:*

$$\begin{aligned} \mathcal{E}_{m,n} &= (\pi^n_{t|t}, f_m) = \int f_m(x)\pi^n_{t|t}(dx) \\ &= \frac{1}{n}\sum_{i=1}^{n} f_m(x^{(i)}) \\ &= \frac{1}{n}\sum_{i=1}^{n} \min([-\log p_t(x^{(i)}], m) \end{aligned} \tag{8}$$

Remark 3 *Note that this evaluation of expectation of a bounded approximation of a function works as a* **Huber M-estimator** *[12]. Huber M-estimator is a term used in robust statistics for a sample mean estimator of expectation which reduces the effect of outlier observations by clipping values of data samples (used to evaluate the sample mean) whose absolute value exceeds an upper bound, m.*

Remark 4 *Note that even when the abnormality occurs, since we do not know the statistics of the changing system, we cannot change the PF parameters to those for the changing system. So for the changing system, $(\pi_{t|t}, f)$ is actually the expected value under the posterior evaluated using wrong filter parameters. But due to asymptotic stability [17], $(\pi_{t|t}, f)$ evaluated using wrong model assumptions converges to the actual one as time goes to infinity. We have shown this rigorously in our recent work, [18] using some results from [17].*

Remark 5 *Also, we would like to clarify that as explained in [11] and other works on particle filtering, the PF is able to track certain changes in the system and/or observation model due to asymptotic stability [17]. As discussed above in remark 4, it is because of this that ELL can be evaluated (with small error) for the changing system using a PF which is optimal for the normal system and can thus be used to detect* *slow changes. But a drastic or abrupt change with n large enough to track only the normal system, will cause the PF to lose track. Now, if the PF does lose track due to a change, tracking error will detect the change while as long as it is able to track the changing system, the evaluated ELL will increase from its normal value, so that ELL can detect such a change.*

6 Abnormal Activity Detection

In a previous work [3], we have defined a PONLD system for modeling the configuration dynamics of a group of interacting point objects which we treat as a moving and deforming shape. The shape deformation model and the motion (scale, rotation) model form the system model and the measurement noise in the point object configuration constitutes the observation noise. We have assumed a stationary shape deformation model, and so we are able to learn a single mean shape for the data. We use the tangent space coordinates w.r.t this mean shape [19] as the shape coordinates. Also, since translation is a linear operation, we use a translation normalized configuration vector as the observation vector (thus avoiding the need to have translation as part of the state vector). We give below the equations used for the observation model and for the system model for shape, scale and rotation.

6.1 Dynamical Model

Observation Model

$$
\begin{aligned}
Y_t &= h(X_t) + w_t, \;\; w_t \sim \mathcal{N}(0, \sigma_{obs}^2 I_{2k}) \quad \textit{where} \\
h(X_t) &= s_t[(1 - v_{tc}^{*} v_{tc})^{1/2} \mu + v_{tc}] e^{-j\theta_t} \qquad (9)
\end{aligned}
$$

Y_t is a complex vector of observations containing the object locations measurements, $h(X_t)$ is a nonlinear function of the state, $X_t = [v_t^T, \theta_t, \mathbf{s}_t]^T$ where v_t is the vector of the tangent coordinates of shape, θ_t is the rotation angle, and s_t is the scale.

System Model

$$
v_t = A v_{t-1} + n_t, \;\; n_t \sim \mathcal{N}(0, \Sigma_n)
$$

$$
\begin{aligned}
s_t &= r_t s_{t-1}, \;\; r_t \sim Rayleigh(\sqrt{2/\pi}) \\
\theta_t &= \theta_{t-1} + u_t, \;\; u_t \sim \text{Unif}(-a, a)
\end{aligned} \tag{10}
$$

The tangent coordinates v_t are assumed to follow a stationary Gauss-Markov model and hence the model parameters A, Σ_n, Σ_v can be learnt using a single training observation sequence. Due to stationarity of v_t, the a-priori pdf of v_t at any time 't' is $p_t(v_t) = \mathcal{N}(0, \Sigma_v)$.

We use in this case only a part of the state vector (the tangent coordinates, v_t) for change detection. The ELL of the tangent coordinates simplifies to $\mathcal{E}_{m,n}(t) = E_{\pi^n_{t|t}}[\min([v_t^T \Sigma_v^{-1} v_t], m)] + C$ where C is a constant. We compare below the performance of ELL with that of tracking error and observation likelihood.

6.2 Simulations and Results

The normal activity (shown in figure 1(a)) is that of a group of people deplaning and walking towards the terminal. Abnormality as shown in 1(b) is one person walking away in a different direction [3]. We have available ground truth values of one realization of the time sequence of configurations of the people (point objects) at each frame which we use to learn the system model. Now the abnormality would be slow or drastic depending on the speed at which the person walks away. We test for different rates of change (slow to drastic) by varying the speed of the person walking away, in simulation. At $t = 5$, the person (shown in figure 1(b)) is made to walk away at 45^o to the X axis with velocity (in both X and Y direction) of $1, 2, 4, 16, 32$ pixels per time step, in 5 different simulations. The average X or Y velocity of any person in a normal sequence is about 1 pixel per time step, hence walk away velcotiy of 1 corresponds to a slow change which can be tracked by the PF while 32 pixels per time step is a very drastic change. Note that observation noise with variance σ^2_{obs} in all directions (spatially white) is added in simulation, to the ground truth data (normal sequence) and the abnormal sequences.

We compare the performances of ELL, tracking error (actually filtering error), $||y_t - E[Y_t|Y_{0:t-1}]||^2 = ||y_t - E_{\pi_{t|t}}[h(X_t)]||^2$ and negative log of likelihood of current observation given past observations, $[-\log Pr(y_t|y_{0:t-1})]$ (Observation likelihood). Figure 2 (a),(b),(c) show the plots of the three statistics as a function of time, for a normal activity sequence, and for walking away velocities of $1, 4, 32$. Observation noise variance is $\sigma^2_{obs} = 9$. As can be seen ELL detects abnormality fastest except for velocity=32 which is very drastic and causes the PF to lose track. We quantify this statement in figure 3.

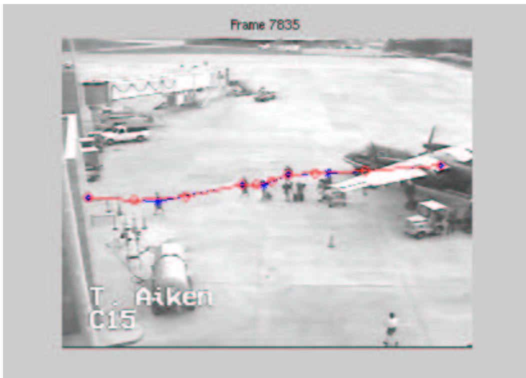
We set as detection threshold for a statistic, its maximum value for a normal test sequence. Using this threshold we plot the detection delay against the rate of change (walk away velocity) for the three statistics. We do this for $\sigma^2_{obs} = 3, 9, 27$ in 3(a),(b),(c). Now detection delay using ELL is least except when the change is very drastic (velocity = 16,32) where tracking error or Obs. likelihood perform better.

The criterion for choosing a change detection statistic and its threshold is to minimize the detection delay for a fixed mean time between false alarms [4]. Since we are assuming unknown change parameters, we plot the average delay in detecting abnormality (averaged over different change rates) against the mean time between false alarms. We do this in figure 4 (a),(b),(c) for $\sigma^2_{obs} = 3, 9, 27$. As can be seen from the figures, ELL has the best average detection delay performance. But we know from the previous figure that tracking error has better performance for drastic abnormalities. Thus as discussed in section 4, we combine ELL and tracking error, i.e. we declare a change if either ELL or tracking error exceed their respective thresholds. To choose an operating point, we vary thresholds for both ELL and tracking error and plot the average detection delay versus mean time between false alarms. Each broken green line plot plots the detection delay for ELL threshold fixed, tracking error threshold varying. As is evident from figure 4, for most values of mean time between false alarm, we can obtain an operating point with this combined strategy that has lower detection delay than either ELL or tracking error alone.

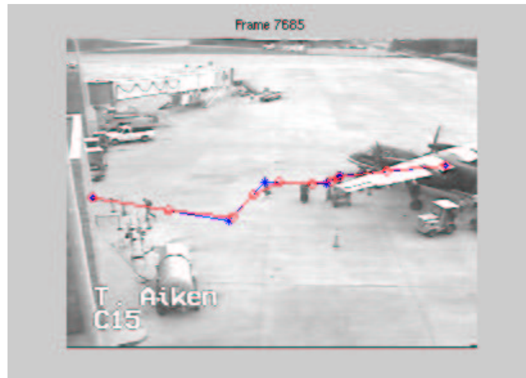
7 Discussion and Future Work

Now [6] defines a CUSUM like likelihood ratio based statistic for change detection. But in our problem, the parameters of the changed system are unknown and so we cannot define the probability of observations under the changed system and hence the likelihood ratio cannot be defined. One could try to adapt the idea to the case of unknown parameters by trying to use $l_t = \max_{1 \le j \le t}[-\log Pr(y_{j:t}|y_{0:j-1}) - T^t_j]$ where $Pr(y_{j:t}|y_{0:j-1})$ is the probability of the observations under the normal system hypothesis and T^t_j is a normalcy threshold for the observation likelihood. Using this kind of a statistic will detect the change time more accurately and have lesser false alarms than just the current 'Obs. likelihood' as defined by us but it is computationally more complex. Also, it is not clear how to set the thresholds T^t_j.

As part of future work, one could attempt to define a CUSUM like statistic for ELL of state given observations i.e. define $\tilde{l}_t = \max_{1 \le j \le t}[-E_{\pi_{j:t|t}} \log p_t(X_{j:t}) - \tilde{T}^t_j]$ and compare its performance with l_t defined above. We can set T^t_j to be the expected value over observation sequences of the ELL at time t,



(a)



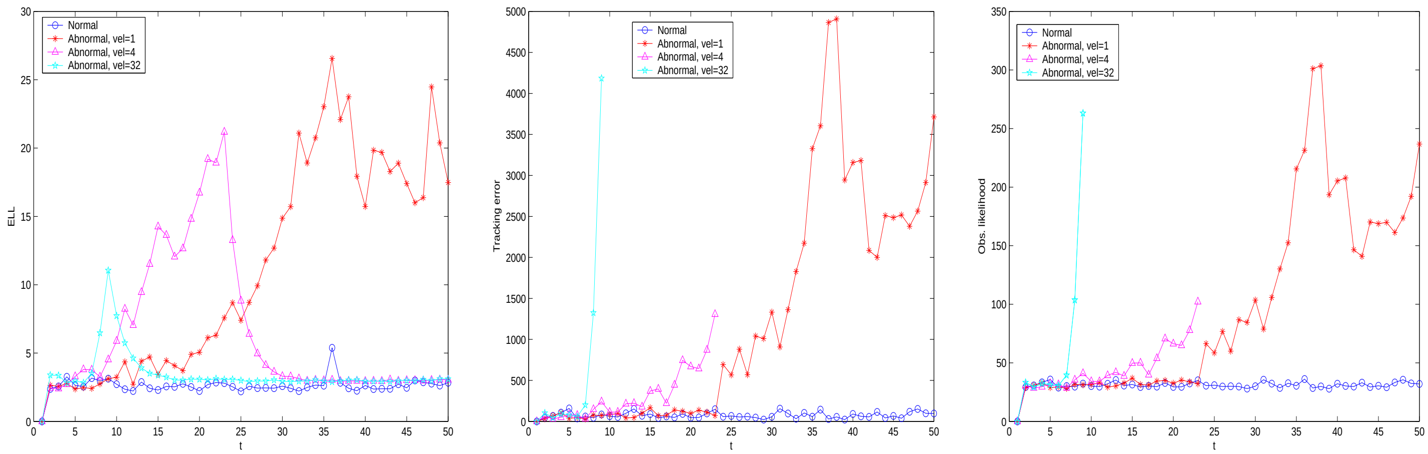
(b)

Figure 1: (a): A 'normal activity' frame with 4 people, (b): Abnormality: One person walking away in a weird direction.

$E_{Y_{0:t}}[ELL(Y_{0:t})]$, which is equal to the expected value under p_t of $-\log p_t(X_{j:t})$ (which is the entropy of $X_{j:t}$). For the case of Gaussian prior this simplifies to a constant times the dimensionality of the data i.e. $(t-j+1)n_x$.

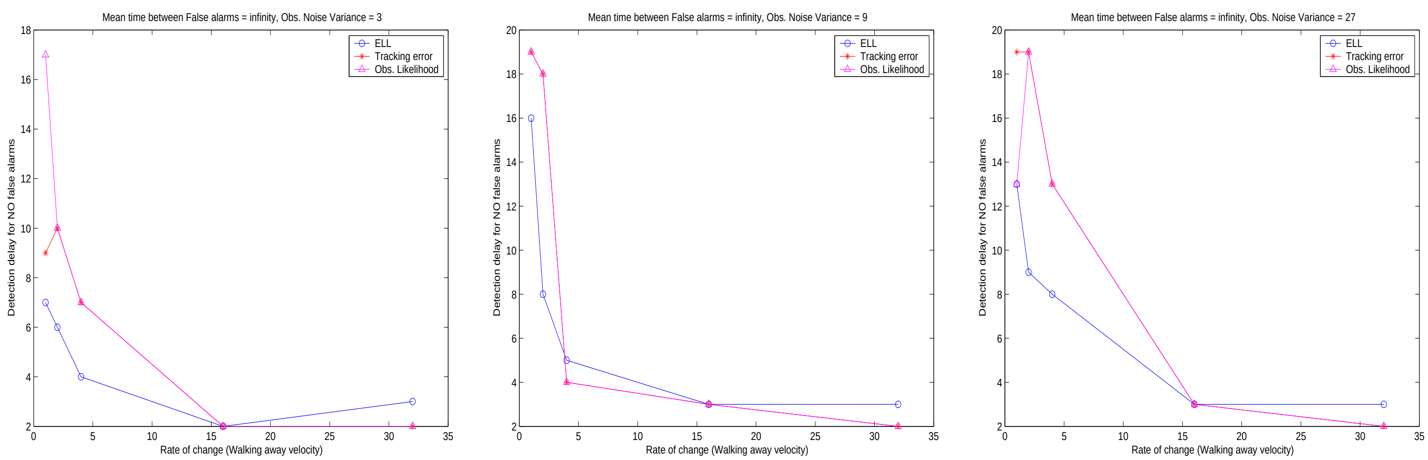
For small observation noises, one could assume the system to be fully observed and replace ELL by log likelihood of $x_{j:t} \approx h^{-1}(y_{j:t})$. It would be interesting to study at what observation noise, the fully observed assumption starts to fail. Finally, we hope to compare performance of the change detection statistics for different kinds of abnormalities in this and other applications and also for non-white and non-Gaussian observation noise. In non-Gaussian noise, one interesting case would be how to distinguish outlier observations caused due to say Cauchy noise from an abnormality. For such a case, a sequence based statistic like $\tilde{l}_t$ or l_t would prove useful. Another interesting problem would be switching between a couple of known modes using a discrete mode variable as in [7, 8] and defining the sequence to be abnormal only if the posterior state distribution is far from the priors corresponding to all the modes.

We have studied the change detection problem in more detail in a recent work [18], which analyzes the errors in approximating the ELL using a PF optimal for the normal system. It also discusses more rigorously, the complementary behavior of ELL and Obs. Likelihood for change detection.



(a) (b) (c)

Figure 2: We show plots of ELL, tracking error and obs. likelihood (in (a),(b),(c)respectively) for normal activity and increasing walk away velocities (abnormal behavior) as a function of time. Abnormality is introduced at $t = 5$ and $\sigma^2_{obs} = 9$.



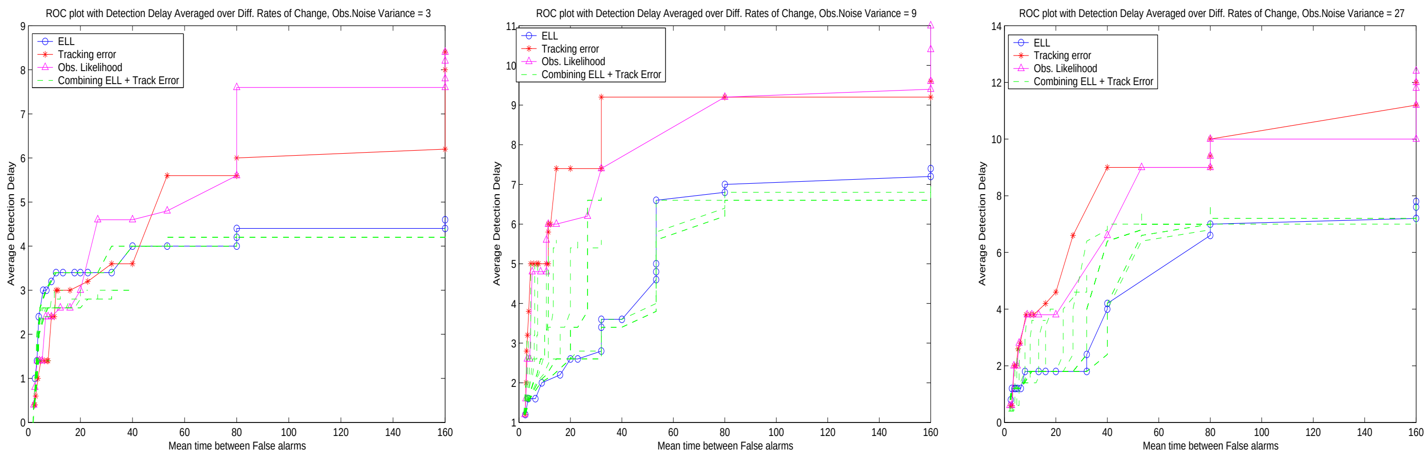
(a) (b) (c)

Figure 3: We plot the detection delay for zero false alarm against the rate of change (walk away velocity) for the three statistics. (a), (b), (c) are for increasing observation noise.

References

[1] A. Doucet, N. deFreitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer, 2001.

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



(a) (b) (c)

Figure 4: We plot the ROC (detection delay averaged over all rates of change versus mean time between false alarms) for the three statistics and the green dotted lines are the ROC for combining ELL and tracking error. (a), (b), (c) are for increasing observation noise.

[3] N. Vaswani, A. RoyChowdhury, and R. Chellappa, "Activity recognition using the dynamics of the configuration of interacting objects," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[4] M. Basseville and I Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice Hall, 1993.

[5] T. Kailath, A.H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, 2000.

[6] B. Azimi-Sadjadi and P.S. Krishnaprasad, "Change detection for nonlinear systems: A particle filtering approach," in *American Control Conference*, 2002.

[7] R. Dearden and D. Clancy, "Particle filters for real-time fault detection in planetary rovers," .

[8] Shaohua Zhou and Rama Chellappa, "Probabilistic human recognition from video," in *European Conference on Computer Vision*, 2002, pp. 681–697.

[9] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, "Novel approach to nonlinear/nongaussian bayesian state estimation," *IEE Proceedings-F (Radar and Signal Processing)*, pp. 140(2):107–113, 1993.

[10] P. Fearnhead, "Sequential monte carlo methods in filter theory," in *PhD Thesis, Merton College, Universoty of OXford*, 1998.

[11] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," *European Conference on Computer Vision*, pp. 343–356, 1996.

[12] G. Casella and R. Berger, *Statistical Inference*, Duxbury Thomson Learning, second edition, 2002.

[13] D. Crisan and A. Doucet, "Convergence of sequential monte carlo methods," in *Technical Report, Cambridge University*, 2000.

[14] D.F. Kerridge, "Inaccuracy and inference," *J. Royal Statist. Society, Ser. B*, vol. 23 1961.

[15] Rudolf Kulhavy, "A geometric approach to statistical estimation," in *IEEE Conference on Decision and Control (CDC)*, Dec. 1995.

[16] H.L. Royden, *Real Analysis*, Prentice Hall, 1995.

[17] LeGland F. and Oudjane N., "Stability and Uniform Approximation of Nonlinear Filters using the Hilbert Metric, and Application to Particle Filters," *Technical report, RR-4215, INRIA*, 2002.

[18] N. Vaswani and R. Chellappa, "Change detection in partially observed nonlinear dynamical systems with change parameters unknown," in *submitted to American Control Conference (ACC)*, 2004.

[19] I.L. Dryden and K.V. Mardia, *Statistical Shape Analysis*, John Wiley and Sons, 1998.