**Fifty-third Annual Allerton Conference**
**Allerton House, UIUC, Illinois, USA**
**September 29 - October 2, 2015**

# Robust Estimation Using Context-Aware Filtering

Radoslav Ivanov[1], Nikolay Atanasov[2], Miroslav Pajic[3], George Pappas[2], and Insup Lee[1]

*Abstract*— **This paper presents the context-aware filter, an estimation technique that incorporates context measurements, in addition to the regular continuous measurements. Context measurements provide binary information about the system's context which is not directly encoded in the state; examples include a robot detecting a nearby building using image processing or a medical device alarming that a vital sign has exceeded a predefined threshold. These measurements can only be received from certain states and can therefore be modeled as a function of the system's current state. We focus on two classes of functions describing the probability of context detection given the current state; these functions capture a wide variety of detections that may occur in practice. We derive the corresponding context-aware filters, a Gaussian Mixture filter and another closed-form filter with a posterior distribution whose moments are derived in the paper. Finally, we evaluate the performance of both classes of functions through simulation of an unmanned ground vehicle.**

## I. Introduction

The problem of state estimation and filtering is pronounced in multiple modern systems where sensors often provide inaccurate readings. For example, in medical applications measurements are frequently missing or wrong as caused by patients moving or clinicians performing other tasks that disrupt the sensors' readings [14]. In robotics, localization is made challenging by imprecise GPS measurements as well as by data association (the problem of determining the correspodence between surrounding objects and received measurements) [3, 17]. These scenarios cause conventional filtering algorithms to perform poorly in such conditions.

To address these problems, in this paper we propose to improve upon current filtering techniques by incorporating context measurements. Context is defined as a system's surrounding environment that is not directly encoded in the state but is correlated with it. Examples of context measurements include a medical device alarming that a vital sign is above a certain threshold or a robot recognizing a nearby building using image processing. What is common among context measurements is that they are binary, i.e., 0 or 1, and they all have a known probability of occurring given the system's state. Thus, the problem considered in this paper is to develop a filter incorporating both continuous and discrete measurements.

Context-aware filtering is similar to target tracking in that measurements may arrive irregularly and from different objects [18]; while general models exist in this setting, they lead to computationally expensive calculations and distributions that cannot be computed in closed form. Another related problem is sensor scheduling where different sensors are used at different times in order to minimize measurement interference or power consumption [15, 29, 31]; however, these techniques only work with continuous measurements and cannot handle the binary readings considered in this paper. A similar problem has been considered in the wireless sensor networks area where it is envisioned that low-cost sensors are deployed over a large area; in such a network, a sensor is only used if the system in consideration (e.g., robot) is nearby [8], which could be considered a context detection in our framework. Finally, prior work also exists on the problem of filtering with intermittent measurements but it also assumes continuous readings [25].

In robotics, there exist multiple works that use context cues for the purposes of localization [5] and mapping by using scene categorization [9, 21, 22] and appearance [6, 30] as well as object class information [2, 3, 4, 10, 24]; however, most techniques assume either only continuous or only semantic (discrete) measurements but do not work with both. The work that is most similar to our paper in its setup and assumptions solves the problem of indoor localization by using both continuous (e.g., range and bearing) and discrete (e.g., type of detected object) measurements [3]. It uses a particle filter to combine the two types of measurements; however, particle filters do not provide any theoretical guarantees for a finite set of particles and suffer from particle-deprivation problems in high-dimensional spaces. Finally, context has also been used in wireless health systems in order to increase certain sensors' weights when performing estimation [19].

In this paper, we aim to solve this problem by deriving exact estimates of the posterior distribution of the state. We note that closed-form expressions for the posterior do not exist for all possible probability of context detection functions. Therefore, we focus on two classes of functions that lead to (near) closed-form solutions and that represent

This material is based on research sponsored by DARPA under agreement number FA8750-12-2-0247. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government. This work was supported by TerraSwarm, one of six centers of STARnet, a Semiconductor Research Corporation program sponsored by MARCO and DARPA. This research was supported in part by Intel-NSF CNS 1505799, NSF CNS-1035715 and by the DGIST Research and Development Program of the Ministry of Science, ICT and Future Planning of Korea (CPS Global Center). Preliminary versions of some of the results in this paper have appeared in the Workshop on Multi-View Geometry in Robotics 2015 [13].

[1]Radoslav Ivanov and Insup Lee are with the Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104, USA {rivanov, lee}@seas.upenn.edu

[2]Nikolay Atanasov and George Pappas are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA {atanasov, pappasg}@seas.upenn.edu

[3]Miroslav Pajic is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA miroslav.pajic@duke.edu



Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:23:46 UTC from IEEE Xplore. Restrictions apply.

a wide variety of detection scenarios observed in practice.

The first class of probability of detection functions are inverse-exponential functions. With this class of functions, the probability of detection is high when the state is close to a certain mean (e.g., the robot is close to a building) and decreases rapidly as the state moves away. We show that this class of functions leads to a closed-form filter with Gaussian Mixtures without any approximations. The second class of functions are sigmoid functions defined as the probit function [20]. The probit function resembles a step function, i.e., for small inputs it is close to 0 but once a threshold is crossed, it increases rapidly and converges to 1. This class of functions capture the threshold medical alarms described above as well as threshold detection scenarios. We derive the mean and covariance matrix of the posterior distribution and argue that a Gaussian distribution with the same moments is a good approximation for it.

Finally, we illustrate the effectiveness of both classes of functions using a case study on an unmanned ground vehicle. In particular, we simulate scenarios in which the vehicle is moving and estimating its state in an environment with very imprecise and biased GPS measurements; at the same time, context measurements are available indicating that speed exceeds a certain threshold or that certain buildings are recognized. We show that the proposed filter using both conventional readings and the binary context measurements outperforms the (extended) Kalman filter which only uses the inaccurate continuous measurements.

The contributions of this paper are the following: 1) a formulation of the context-aware filtering problem for linear systems; 2) the development of a Guassian-Mixture-based filter and a sigmoid-based filter using the two proposed classes of probability of detection functions, respectively; 3) an illustration of the effectiveness of context-aware filtering using a case study on an unmanned ground vehicle.

The rest of this paper is organized as follows. Section II formulates the context-aware filtering problem. Section III outlines a general approach and explains its deficiencies when the probability of detection function is unknown. Sections IV and V focus on the inverse-exponential and sigmoid function classes, respectively, and develop the resulting filters. In Section VI we provide the case study evaluation. Finally, Section VII concludes the paper.

## II. Problem Formulation

The context-aware state estimation and filtering problem is as follows. Consider a linear discrete-time system of the form

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \tag{1}$$

where $x \in \mathbb{R}^n$ is the system state, $u \in \mathbb{R}^p$ is the applied input, $w \sim \mathcal{N}(0, Q)$ is Gaussian process noise, and $A$ and $B$ are matrices of appropriate dimensions.

The system has two kinds of sensors available to it: state and context. State sensors measure (subsets of) the state directly. We assume that there is a known linear observation model for state sensors of the form

$$y_k = C_k x_k + v_k, \tag{2}$$

where we denote state sensors' measurements by $y_k \in \mathbb{R}^m$, $v_k \sim \mathcal{N}(0, R)$ is Gaussian measurement noise, and matrix $C$ has appropriate dimensions.

Context sensors, on the other hand, do not measure the system's state but rather provide binary information about its context. We define context as a finite set $\mathcal{C} = \{c_1, \ldots, c_N\}$, where each $c_i$ is a context element that can be detected by a context sensor from certain states; example context elements include nearby buildings with known positions on a map or a vital sign exceeding a certain predefined threshold. For each $i$, a measurement $b_k^i$ is received that is 1 if $c_i$ is detected and 0 otherwise. We do not assume knowledge about a context element, other than the probability of detection given a state, denoted by $p_d(c_i \mid x)$, i.e.,

$$b_k^i = \begin{cases} 1 & w.p. \quad p_d(c_i \mid x) \\ 0 & w.p. \quad 1 - p_d(c_i \mid x), \end{cases} \tag{3}$$

where $p_d$ is a function of the system state. Thus, the set of context measurements is $b_k \in \{0,1\}^N$. It is assumed that, conditioned on the state, context measurements are mutually independent and are also independent from state measurements.

**Problem.** *Given the system defined in (1)-(3) and a prior pdf $p_{k|k}(x) = p(x \mid u_{0:k-1}, y_{0:k}, b_{0:k})$ the goal is to compute the posterior density*

$$p_{k+1|k+1}(x) := p(x \mid u_{0:k}, y_{0:k+1}, b_{0:k+1}),$$

*describing the system's state given all available measurements and inputs.*

## III. General Approach

The problem defined in Section II naturally lends itself to a Bayesian approach with a predict and update phase of the form

$$\textbf{Predict: } p_{k+1|k}(x) = \int p_f(x \mid z, u_k) p_{k|k}(z) dz, \tag{4}$$

$$\textbf{Update: } p_{k+1|k+1}(x) = \eta_{k+1} p_o(y_{k+1}, b_{k+1} \mid x) p_{k+1|k}(x),$$

where $p_f(x_{k+1} \mid x_k, u_k)$ is the conditional pdf of the state at time $k+1$ given the state and input at time $k$, $p_o(y_{k+1}, b_{k+1} \mid x_{k+1})$ is the joint pdf of all available measurements (state and context) given the state and $\eta_{k+1}$ is a normalization constant [28].

While (4) provides a compact representation of the filtering problem, in general it is impossible to obtain a closed-form expression for the densities. The notable exception is the linear Gaussian case, which leads to the Kalman filter [16]. Multiple extensions to the Kalman filter have been proposed, e.g., by computing first-order Taylor series approximations of nonlinear densities; such techniques are known as the extended Kalman filter (EKF) or other linear regression Kalman filters [11]. These techniques cannot be directly applied to the problem considered in this paper since context measurements introduce discontinuities in the pdf's that cannot be handled by current algorithms.

One approach when such discontinuities exist is to employ Markov Chain Monte Carlo approximations (e.g., particle filters [28]) or other nonparametric sampling techniques such



Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:23:46 UTC from IEEE Xplore. Restrictions apply.

as Gaussian Processes approximations [11]. However, while these techniques have nice convergence properties when the number of particles gets large, they do not provide theoretical guarantees for any finite set of samples. Therefore, in this paper we focus on two specific probability of context detection functions (i.e., $p_d(c_i \mid x)$ in (3)) that lead to closed-form filters. We argue that each of these functions captures a sufficiently large class of scenarios so as to be useful in a lot of modern systems. We describe each of these functions, and the resulting filters, in the following two sections.

## IV. Inverse-Exponential Functions

The first class of probability of context detection functions considered in this paper are inverse-exponential functions.

**Assumption.** *Suppose the probability of context detection functions are inverse-exponential functions that are defined as follows:*

$$p_d(c_i \mid x_k) = e^{-\frac{1}{2}(G_i x_k - \theta_i)^T V_i^{-1} (G_i x_k - \theta_i)}, \tag{5}$$

*which are parameterized by $\theta_i \in \mathbb{R}^q$ and $V_i \in \mathbb{R}^{q \times q}$, the analogues of a mean and a covariance in a multivariate Gaussian pdf, and $G_i \in \mathbb{R}^{q \times n}$, which can be thought of as a selection matrix when $q < n$. This probability is 1 when $G_i x_k = \theta_i$ and approaches 0 when $G_i x_k - \theta_i$ gets large.*

We argue that these functions capture many scenarios observed in reality. For example, in the case of detecting a known nearby building using image processing, the probability of detecting the building is high when the camera is closer to the building and decreases rapidly as it moves away [3]. In addition, it might be easier to detect the building from certain angles than it is from others, and this is also captured by the above formulation.

Having fixed (5) as the probability of context detection, with $\theta_i$ and $V_i$ known for each context element (or potentially learned from data), we show how to derive the recursive context-aware filter. Note that, after including the binary measurements, the filter has a closed-form solution for Gaussian Mixtures (GMs), i.e., if the prior is a GM, then so is the posterior. A GM is a distribution whose pdf is defined as a weighted sum of Gaussian pdfs:

$$g(x) = \sum_{i=1}^{M} w_i \phi(x; \mu_i, \Sigma_i), \tag{6}$$

where $\phi(x; \mu_i, \Sigma_i)$ is the pdf of a Gaussian distribution with mean $\mu_i$ and covariance matrix $\Sigma_i$, and $w_i$ are weights such that $\sum_{i=1}^{M} w_i = 1$. GMs have two properties that make them attractive for modeling multimodal distributions. First of all, they are linear combinations of Gaussian pdfs, thus a recursive filter using a GM can be developed with a bank of Kalman filters, one for each element in the GM. In addition, with a sufficient number of elements, a GM can be used to approximate any continuous pdf [11]. Finally, it is important to note that a Gaussian is a special case of a GM, hence it is possible to initialize the filter with a Gaussian prior on $x_0$.

**Proposition 1.** *Consider a system with a linear motion model*

$$x_{k+1} = A_k x_k + B_k u_k + w_k,$$

*linear state observation model*

$$y_k = C_k x_k + v_k,$$

*and context observation of the form*

$$p_d(c_i \mid x_k) = e^{-\frac{1}{2}(G_i x_k - \theta_i)^T V_i^{-1} (G_i x_k - \theta_i)}.$$

*Assuming that the state prior $p_{k|k}$ is a Gaussian Mixture, then the predicted and updated pdf's, $p_{k+1|k}$ and $p_{k+1|k+1}$ respectively, are also Gaussian Mixtures **without any approximation**.*

*Proof.* Note that, unlike the conventional Kalman filter that has a predict and an update stage, the proposed filter has three steps: prediction, continuous update and discrete update. There is also an optional mixture reduction step discussed at the end of the section.

### A. Predict

For the predict stage, we note that

$$\begin{aligned} p_{k+1|k}(x) &= \sum_{i=1}^{M} w_i \int \phi(x; A_k z + B_k u_k, Q) \phi(z; \mu_i, \Sigma_i) dz \\ &= \sum_{i=1}^{M} w_i \phi(x; A_k \mu_i + B_k u_k, A_k \Sigma_i A_k^T + Q) \\ &= \sum_{i=1}^{M} w_i \phi(x; \mu_i^p, \Sigma_i^p), \end{aligned}$$

which is the usual form of the Kalman filter predict equations (e.g., see [16]). The resulting distribution is again a GM.

### B. Continuous Update

As described above, we perform the update separately for state (continuous) and context (discrete) sensors. Upon receiving a measurement $y_{k+1}$, the continuous update is:

$$\begin{aligned} p^c_{k+1|k+1}(x) &= \frac{p(y_{k+1} \mid x) p_{k+1|k}(x)}{\int p(y_{k+1} \mid z) p_{k+1|k}(z) dz} \\ &= \frac{\phi(y_{k+1}; C_k x, R) \sum_{i=1}^{M} w_i \phi(x; \mu_i^p, \Sigma_i^p)}{\int \phi(y_{k+1}; C_k z, R) \sum_{j=1}^{M} w_j \phi(z; \mu_j^p, \Sigma_j^p) dz} \\ &= \sum_{i=1}^{M} \left( \frac{w_i \gamma_i^c}{\alpha^c} \right) \frac{\phi(y_{k+1}; C_k x, R) \phi(x; \mu_i^p, \Sigma_i^p)}{\int \phi(y_{k+1}; C_k z, R) \phi(z; \mu_i^p, \Sigma_i^p) dz} \\ &= \sum_{i=1}^{M} w_i^c \phi(x; \mu_i^c, \Sigma_i^c), \end{aligned}$$



Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:23:46 UTC from IEEE Xplore. Restrictions apply.

where

$$\begin{aligned}
\alpha^c &:= \sum_{i=1}^{M} w_i \gamma_i^c \\
\gamma_j^c &:= \int \phi(y_{k+1}; C_k z, R)\phi(z; \mu_j^p, \Sigma_j^p)dz \\
&= \phi(y_{k+1}; C_k \mu_j^p, C_k \Sigma_j^p C_k^T + R) \\
\mu_j^c &:= \mu_j^p + K_j^c(y_{k+1} - C_k \mu_j^p) \\
\Sigma_j^c &:= (I - K_j^c C_k)\Sigma_j^p \\
K_j^c &:= \Sigma_j^p C_k^T (C_k \Sigma_j^p C_k^T + R)^{-1}.
\end{aligned}$$

Note that the posterior distribution is also a GM with the same number of elements but with possibly rescaled weights.

### C. Discrete Update

For the discrete update, to simplify notation we assume that there is only one context element, with corresponding parameters $\theta$, $V$ and $G$, such that $b_{k+1} \in \{0, 1\}$, though the approach can be straightforwardly modified to include multiple binary measurements through repeated updates.

First note that the posterior distribution depends on whether $b_{k+1}$ is 0 or 1 as the probabilities of getting either one are different. Consider first the case when $b_{k+1} = 1$:

$$\begin{aligned}
p_{k+1|k+1}(x) &= \frac{p(b_{k+1} = 1 \mid x) p^c_{k+1|k+1}(x)}{\int p(b_{k+1} = 1 \mid z) p^c_{k+1|k+1}(z) dz} \\
&= \frac{\phi(\theta; Gx, V) \sum_{i=1}^{M} w_i^c \phi(x; \mu_i^c, \Sigma_i^c)}{\int \phi(\theta; Gz, V) \sum_{j=1}^{M} w_j^c \phi(z; \mu_j^c, \Sigma_j^c) dz} \\
&= \sum_{i=1}^{M} \left( \frac{w_i^c \gamma_i^d}{\alpha^d} \right) \frac{\phi(\theta; Gx, V)\phi(x; \mu_i^c, \Sigma_i^c)}{\int \phi(\theta; Gz, V)\phi(z; \mu_i^c, \Sigma_i^c) dz} \\
&= \sum_{i=1}^{M} \left( \frac{w_i^c \gamma_i^d}{\alpha^d} \right) \phi(x; \mu_i^d, \Sigma_i^d),
\end{aligned}$$

where $\alpha^d$, $\gamma_j^d$, $\mu_j^d$, $\Sigma_j^d$ and $K_j^d$ are defined similar to their continuous analogues.

Finally, when $b_{k+1} = 0$, the update becomes

$$\begin{aligned}
p_{k+1|k+1}(x) &= \frac{(1 - p(b_{k+1} = 1 \mid x))\, p^c_{k+1|k+1}(x)}{\int (1 - p(b_{k+1} = 1 \mid z))\, p^c_{k+1|k+1}(z) dz} \\
&= \sum_{i=1}^{M} \frac{w_i^c\, (1 - p(b_{k+1} = 1 \mid x))\, \phi(x; \mu_i^c, \Sigma_i^c)}{1 - \sum_{j=1}^{M} w_j^c \int p(b_{k+1} = 1 \mid z) \phi(z; \mu_i^c, \Sigma_i^c) dz} \\
&= \sum_{i=1}^{M} \frac{w_i^c}{1 - \sum_{j=1}^{M} w_j^c \beta_j} \phi(x; \mu_i^c, \Sigma_i^c) + \\
&\quad + \frac{-w_i^c \beta_i}{1 - \sum_{j=1}^{M} w_j^c \beta_j} \phi(x; \mu_i^d, \Sigma_i^d)
\end{aligned}$$

where $\beta_j := \gamma_j^d \sqrt{(2\pi)^q \det(V)}$. □

Thus, we have inductively shown that for the probability of detection function considered in this paper, the localization filter can be computed in closed form and results in a GM distribution of the posterior. Note that the number of elements in the GM doubles every time $b_k = 0$, thus an additional step may be necessary in order to bound the number of elements.

### D. Mixture Reduction

The proof of Proposition 1 provides an exact form for the posterior. However, the number of elements in the GM doubles every time a measurement of 0 is received; hence, this number may increase exponentially over time. Many approaches for reducing the number of elements have been proposed in the literature, ranging from keeping the elements with highest weights to merging or discarding elements based on certain notions of distance between them [26]. Note that most available techniques assume weights are positive, yet the GM developed in this paper may have negative weights as well. In such cases, one may use a Gibbs Sampler [27] in order to reduce the size of the GM. A Gibbs Sampler draws random samples from the distribution and can approximate it with a GM with a desired number of elements.

## V. Sigmoid Functions

Note that while the inverse-exponential functions defined in Section IV are general and capture a wide variety of scenarios, they still do not cover all interesting applications. In particular, a major limitation is that they are symmetric around their "mean", i.e., $\theta_i$. However, there are many situations where non-symmetric detections may occur. In the medical domain, for example, it is common to set threshold alarms, i.e., an alarm is raised when a vital sign exceeds a certain threshold (e.g., blood pressure is higher than 160 millimeters of mercury). Similarly, a building may only be recognizable from one angle (e.g., a coffee shop that is only visible from the front of a building).

Thus in this section, we investigate another class of probability of detection functions that capture these non-symmetric cases.

**Assumption.** *Suppose the probability of context detection functions are sigmoid functions that are defined as the probit logistic function [20]:*

$$p_d(c_i \mid x_k) = \Phi(v_i^T x_k + a_i), \tag{7}$$

*where $\Phi$ is the cumulative distribution function of the standard Normal distribution, $v_i \in \mathbb{R}^n$ is a vector of known parameters, and $a_i \in \mathbb{R}$ is a known parameter offset.*

Note that the inner function in (7) defines a halfspace determined by the values of $v_i$ and $a_i$. In the one-dimensional case, this function has a very similar shape to the classical sigmoid function: $f(x) = 1/(1+e^{-x})$. Thus, the probability of detection is close to 0 for values of $x$ far from the "threshold" but once the "threshold" is crossed, the probability of detection increases rapidly and approaches 1. This function is very suitable for modeling the non-symmetric scenarios described above as it is expected that once a signal exceeds a certain threshold, even inaccurate sensors will be able to detect the event and raise an alarm.

Developing an exact filter with the probability of detection defined in (7) is not straightforward, however, as the posterior distribution, after incorporating binary measurements, is not the same as the prior (even if the prior is a Gaussian). As we argue below, however, a Gaussian distribution with the



Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:23:46 UTC from IEEE Xplore. Restrictions apply.

same mean and covariance matrix is a good approximation for the resulting posterior distribution.

In the following subsections we present the phases of the sigmoid-based filter, similar to the GM-based one. In this case we assume the prior $p_{k|k}$ is a **single Gaussian distribution with mean $\mu$ and covariance matrix $\Sigma$.**

### A. Predict

The predict stage in this case is the classical Kalman filter update:

$$\begin{aligned} p_{k+1|k}(x) &= \int \phi(x; A_k z + B_k u_k, Q)\phi(z; \mu, \Sigma)dz \\ &= \phi(x; A_k\mu + B_k u_k, A_k \Sigma A_k^T + Q) \\ &= \phi(x; \mu^p, \Sigma^p). \end{aligned}$$

### B. Continuous Update

The continuous update is also the same as in the Kalman filter:

$$\begin{aligned} p^c_{k+1|k+1}(x) &= \frac{p(y_{k+1} \mid x) p_{k+1|k}(x)}{\int p(y_{k+1} \mid z) p_{k+1|k}(z) dz} \\ &= \frac{\phi(y_{k+1}; C_k x, R)\phi(x; \mu^p, \Sigma^p)}{\int \phi(y_{k+1}; C_k z, R)\phi(z; \mu^p, \Sigma^p)dz} \\ &= \phi(x; \mu^c, \Sigma^c), \end{aligned}$$

where

$$\begin{aligned} \mu^c &:= \mu^p + K^c(y_{k+1} - C_k \mu^p) \\ \Sigma^c &:= (I - K^c C_k)\Sigma^p \\ K^c &:= \Sigma^p C_k^T (C_k \Sigma^p C_k^T + R)^{-1}. \end{aligned}$$

### C. Discrete Update

Similar to the discrete update in Section IV, to simplify notation we assume a single context measurement, $b_k$, with corresponding parameters $v$ and $a$; the approach can be straightforwardly extended to multiple measurements through repeated updates.

**Proposition 2.** *When $b_k = 1$, the discrete update is as follows:*

$$p_{k+1|k+1}(x) = \frac{\Phi(v^T x + a)\phi(x; \mu^c, \Sigma^c)}{Z_1}, \tag{8}$$

*where*

$$Z_1 = \Phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}\right).$$

*Proof.* First note that when $b_k = 1$

$$\begin{aligned} p_{k+1|k+1}(x) &= \frac{p(b_k = 1 \mid x)\phi(x; \mu^c, \Sigma^c)}{\int p(b_k = 1 \mid x')\phi(x'; \mu^c, \Sigma^c)dx'} \\ &= \frac{\Phi(v^T x + a)\phi(x; \mu^c, \Sigma^c)}{Z_1}, \end{aligned}$$

where

$$Z_1 = \int \Phi(v^T x' + a)\phi(x'; \mu^c, \Sigma^c)dx'.$$

The derivation for $Z_1$ when $b_k = 1$ is carried out as follows:

$$\begin{aligned} Z_1 &= \int \Phi(v^T x' + a)\phi(x'; \mu^c, \Sigma^c)dx' = \mathbb{E}_x\left[\Phi(v^T x + a)\right] \\ &= \mathbb{E}_x\left[\mathbb{P}(y \le v^T x + a)\right] = \mathbb{E}_{(x,y)}\left[\mathbb{1}_{y \le v^T x + a}\right] \\ &= \mathbb{P}(v^T x + a - y \ge 0) \\ &= \mathbb{P}\left(v^T \mu^c + a + z\sqrt{v^T \Sigma^c v + 1} \ge 0\right) \\ &= \mathbb{P}\left(z \ge \frac{-v^T \mu^c - a}{\sqrt{v^T \Sigma^c v + 1}}\right) \\ &= 1 - \Phi\left(\frac{-v^T \mu^c - a}{\sqrt{v^T \Sigma^c v + 1}}\right) = \Phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}\right) \end{aligned}$$

where $y$ and $z$ are standard Normal random variables independent of each other and of $x$. □

**Proposition 3.** *When $b_k = 0$, the discrete update is as follows:*

$$p_{k+1|k+1}(x) = \frac{\Phi(-v^T x - a)\phi(x; \mu^c, \Sigma^c)}{Z_0}, \tag{9}$$

*where*

$$Z_0 = \Phi\left(\frac{-v^T \mu^c - a}{\sqrt{v^T \Sigma^c v + 1}}\right).$$

*Proof.* Note that, since $\Phi$ has rotational symmetry, $p(b_k = 0 \mid x) = 1 - \Phi(v^T x + a) = \Phi(-v^T x - a)$. The rest of the proof is the same as the proof of Proposition 2. □

**Approximation.** *We approximate the posterior distribution in (8) and (9) with a Gaussian distribution with the same mean and covariance matrix.*

Note that the posterior distribution after incorporating context measurements is no longer Gaussian. However, a Gaussian still seems to be a good approximation for (8) and (9); while it is challenging to show that the distribution in (8) and (9) is unimodal, simulations and similar results lead us to believe so. In particular, Ibragimov [12] proved that the convolution of a log-concave and a unimodal random variable is also a unimodal random variable. Equations (8) and (9) do not involve convolution but have the same form of multiplication of two densities; additionally, $\Phi(v^T x + a)$ is log-concave as shown in Chapter 2.5 of [7] and $\phi(x; \mu^c, \Sigma^c)$ is unimodal since it is the pdf of a Gaussian distribution. Thus, we approximate the posteriors in (8) and (9) with a Gaussian with the same mean and covariance matrix and leave the approximation error analysis for future work.

**Proposition 4.** *The mean of the distribution in (8) is:*

$$\mu_1 = \mu^c + \Sigma^c v \frac{\phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}; 0, 1\right)}{Z_1 \sqrt{v^T \Sigma^c v + 1}}. \tag{10}$$

*The covariance matrix of the distribution in (8) is:*

$$\begin{aligned} \Sigma_1 = \Sigma^c &+ \mu_1(\mu^c)^T + \mu^c \mu_1^T - \mu^c(\mu^c)^T \\ &- \Sigma^c v v^T \Sigma^c \frac{\phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}; 0, 1\right)(v^T \mu^c + a)}{Z_1 (v^T \Sigma^c v + 1)^{3/2}} - \mu_1 \mu_1^T. \end{aligned}$$



Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:23:46 UTC from IEEE Xplore. Restrictions apply.

*Proof.* First note that

$$\mu_1 = \int x' \frac{\Phi(v^T x' + a)\phi(x'; \mu^c, \Sigma^c)}{Z_1} dx'.$$

One way to compute the mean in closed-form is, similar to the derivation in Chapter 3.9 in [23], by first computing the gradient with respect to $\mu^c$ of the following two equivalent expressions for $Z_1$:

$$\int \Phi(v^T x' + a)\phi(x'; \mu^c, \Sigma^c) dx' = \Phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}\right), \tag{11}$$

i.e.,

$$\frac{\partial Z_1}{\partial \mu^c} = \int (\Sigma^c)^{-1}(x' - \mu^c)\Phi(v^T x' + a)\phi(x'; \mu^c, \Sigma^c) dx'$$
$$= v \frac{\phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}; 0, 1\right)}{\sqrt{v^T \Sigma^c v + 1}},$$

where we used the fact that $\partial \Phi(x)/\partial \mu = \phi(x)\partial(x)/\partial \mu$. Note that the first term in the integral on the left-hand side is $Z_1(\Sigma^c)^{-1}\mu_1$. The second term is $Z_1(\Sigma^c)^{-1}\mu^c$. Therefore, by rearranging terms, we get

$$Z_1(\Sigma^c)^{-1}\mu_1 = Z_1(\Sigma^c)^{-1}\mu^c + v \frac{\phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}; 0, 1\right)}{\sqrt{v^T \Sigma^c v + 1}}.$$

Thus, we arrive at the final expression:

$$\mu_1 = \mu^c + \Sigma^c v \frac{\phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}; 0, 1\right)}{Z_1 \sqrt{v^T \Sigma^c v + 1}}.$$

The expression for the covariance matrix is:

$$\Sigma_1 = \hat{\Sigma}_1 - \mu_1 \mu_1^T, \tag{12}$$

where

$$\hat{\Sigma}_1 = \int x' x'^T \frac{\Phi(v^T x' + a)\phi(x'; \mu^c, \Sigma^c)}{Z_1} dx'.$$

$\hat{\Sigma}_1$ can be computed in closed-form similar to the mean, by computing the Hessians with respect to $\mu^c$ of both sides of (11):

$$\int (\Sigma^c)^{-1}(x' - \mu^c)(x' - \mu^c)^T (\Sigma^c)^{-1} \Phi(v^T x' + a)\phi(x'; \mu^c, \Sigma^c) dx'$$
$$- \int (\Sigma^c)^{-1} \Phi(v^T x' + a)\phi(x'; \mu^c, \Sigma^c) dx'$$
$$= -vv^T \frac{\phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}; 0, 1\right)(v^T \mu^c + a)}{(v^T \Sigma^c v + 1)^{3/2}}.$$

Note that one of the terms in the integral on the right-hand side is $Z_1(\Sigma^c)^{-1}\hat{\Sigma}_1(\Sigma^c)^{-1}$. Therefore, we rearrange terms and divide by $Z_1$ to obtain the following:

$$(\Sigma^c)^{-1}\hat{\Sigma}_1(\Sigma^c)^{-1} = (\Sigma^c)^{-1} + (\Sigma^c)^{-1}\mu_1(\mu^c)^T(\Sigma^c)^{-1}$$
$$+ (\Sigma^c)^{-1}\mu^c\mu_1^T(\Sigma^c)^{-1} - (\Sigma^c)^{-1}\mu^c(\mu^c)^T(\Sigma^c)^{-1}$$
$$- vv^T \frac{\phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}; 0, 1\right)(v^T \mu^c + a)}{Z_1(v^T \Sigma^c v + 1)^{3/2}}.$$



Fig. 1. LandShark vehicle [1].

Finally, we arrive at the expression for $\hat{\Sigma}_1$:

$$\hat{\Sigma}_1 = \Sigma^c + \mu_1(\mu^c)^T + \mu^c\mu_1^T$$
$$- \mu^c(\mu^c)^T - \Sigma^c vv^T\Sigma^c \frac{\phi\left(\frac{v^T \mu^c + a}{\sqrt{v^T \Sigma^c v + 1}}; 0, 1\right)(v^T \mu^c + a)}{Z_1(v^T \Sigma^c v + 1)^{3/2}}.$$

Thus, the covariance matrix can be computed by plugging in the expression for $\hat{\Sigma}_1$ in (12). □

**Proposition 5.** *The mean of the distribution in (9) is:*

$$\mu_0 = \mu^c - \Sigma^c v \frac{\phi\left(\frac{-v^T \mu^c - a}{\sqrt{v^T \Sigma^c v + 1}}; 0, 1\right)}{Z_0 \sqrt{v^T \Sigma^c v + 1}}. \tag{13}$$

*The covariance matrix of the distribution in (9) is:*

$$\Sigma_0 = \Sigma^c + \mu_0(\mu^c)^T + \mu^c\mu_0^T - \mu^c(\mu^c)^T$$
$$+ \Sigma^c vv^T\Sigma^c \frac{\phi\left(\frac{-v^T \mu^c - a}{\sqrt{v^T \Sigma^c v + 1}}; 0, 1\right)(v^T \mu^c + a)}{Z_0(v^T \Sigma^c v + 1)^{3/2}} - \mu_0\mu_0^T.$$

*Proof.* The proof is similar to that of Proposition 4 and is omitted in the interest of space. □

## VI. Case Study

In this section we provide evaluations of each of the two filters developed in the paper, beginning with the GM-based filter. In both case studies, we use the LandShark [1] robot (Figure 1) as our platform; the LandShark is an unmanned ground vehicle that is used to perform critical missions on enemy territory, e.g., save injured people. It is equipped with multiple sensors to estimate position and velocity but some of them may perform poorly under certain conditions (e.g., GPS in an urban environment [17]).

### A. Localization Using Inverse-Exponential Functions

To evaluate the effectiveness of the GM-based filter, we simulate the following scenario. The LandShark is moving in an urban environment and trying to visit different waypoints as part of its mission. The robot is using GPS measurements to localize; however, they have a large variance and a bias to the North, thus making localization challenging. To alleviate this problem, the LandShark also has a camera that it uses to recognize nearby buildings using image processing; these recognitions constitute the context measurements.

The entire trajectory, including the city's map, is given in Figure 2. Note that the LandShark has a differential-drive model, hence every turn results in nonlinear dynamics; thus, a linearization is necessary in such cases, as in a typical EKF.



Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:23:46 UTC from IEEE Xplore. Restrictions apply.

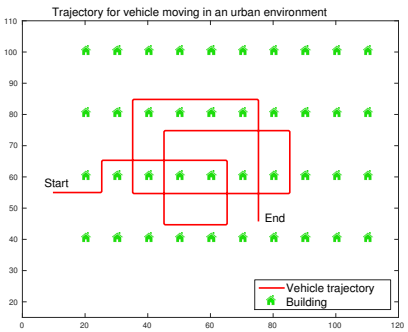



Fig. 2. Entire LandShark trajectory.

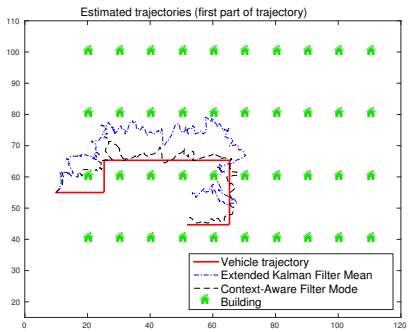



Fig. 3. Estimated trajectories.

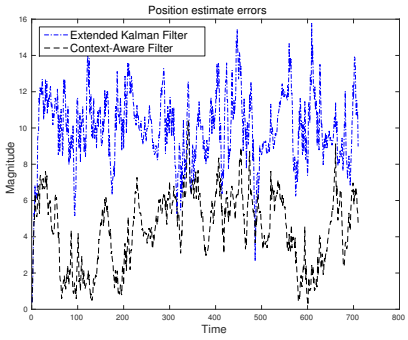



Fig. 4. Position errors by each filter.

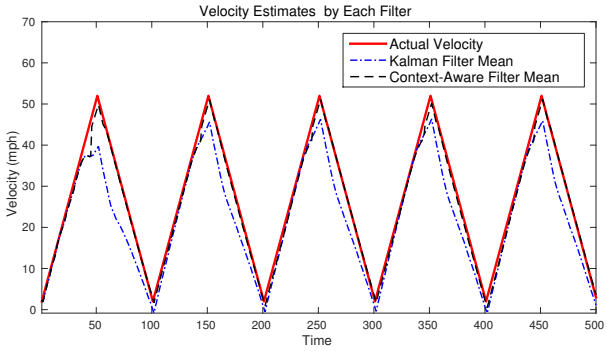



Fig. 5. Velocity estimates by each filter.

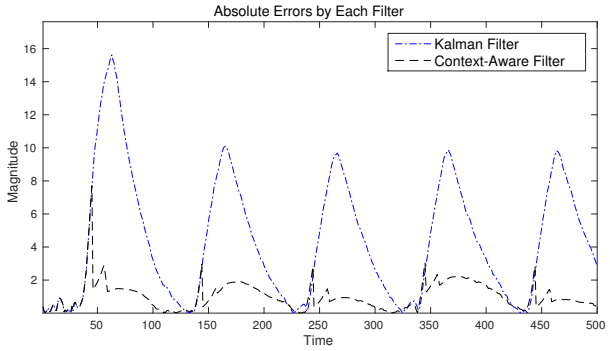



Fig. 6. Absolute errors by each filter.

To evaluate the performance of the GM-based filter, we compare it with a regular EKF that uses only the continuous GPS measurements. Figure 3 shows the path estimates of each of the two filters for the first part of the trajectory. For the EKF, the estimate is the mean of the posterior Gaussian distribution at each time step; for the GM-based filter, the estimate is the mode of the distribution (in this application the mode is selected as the element with highest weight in the mixture). As can be seen in the Figure, the context-aware filter estimates are consistently better and are much more robust to the large variance of the GPS measurements, whereas the EKF's estimates have great jumps when inaccurate measurements arrive.

Finally, Figure 4 provides another confirmation of the better performance of the context-aware filter. It shows each filter's position estimate errors for the LandShark's entire trajectory. The context-aware filter's errors are almost invariably lower than those of the EKF and do not tend to vary greatly from one round to the next. Thus, we conclude that the context-aware filter with inverse-exponential functions outperforms current filtering techniques using only continuous measurements.

### B. Velocity Estimation Using Sigmoid Functions

To illustrate the effectiveness of the sigmoid detection functions, we simulate a modified version of the previous scenario. The LandShark is moving in a straight line in an urban environment, accelerating to a target velocity and then slowing down as it approaches intersections. In this scenario the LandShark has to obtain accurate velocity estimates in order to avoid collisions at intersections while moving as quickly as possible. To estimate velocity, it uses GPS measurements that have a negative bias at high speeds. In addition, the LandShark has access to one context measurement; it is raised by a sensor measuring air resistance at the front of the vehicle. While mapping resistance to speed is not straightforward, it is possible to establish whether the vehicle is moving beyond a certain velocity threshold. Thus, the context measurement is an alarm indicating that the LandShark is approaching its target velocity.

To evaluate the performance of the context-aware filter, we compare it with a Kalman filter that is only using the continuous measurements. The velocity estimates for each filter are shown in Figure 5. As one can see in the Figure, both filters are fairly accurate when GPS readings are unbiased at low speeds; as soon as the bias is introduced, however, the Kalman filter's estimates are greatly affected whereas the context-aware filter remains robust to the bad measurements. Finally, Figure 6 displays the absolute errors incurred by each filter; the Kalman filter's average error is 4.41, whereas the context-aware filter achieves an average error of 1.03. Therefore, this case study also shows that context-aware filters are a promising direction for future work in scenarios where sensor errors might occur.

## VII. Conclusion

In this paper, we presented the context-aware filter that incorporates binary context measurements in addition to regular continous measurements. We focused on two classes of probability of detection functions that capture a wide variety of detection events that may occur in practice. We derived the resulting filters, a closed-form Gaussian Mixture filter and an approximation Gaussian filter with the same first two moments as the true posterior distribution. Finally,



Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:23:46 UTC from IEEE Xplore. Restrictions apply.

we presented a case-study evaluation of each of the two functions using simulations on an unmanned ground vehicle. As part of future work, we will prove the unimodality of the posterior distribution in the sigmoid-based filter and will evaluate the context-aware filter on real data from medical and robotics applications.

## REFERENCES


[1] The LandShark. http://blackirobotics.com/LandShark_UGV_UC0M.html.

[2] R. Anati, D. Scaramuzza, K. Derpanis, and K. Daniilidis. Robot localization using soft object detection. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 4992–4999, 2012.

[3] N. Atanasov, M. Zhu, K. Daniilidis, and G. Pappas. Semantic localization via the matrix permanent. In *Robotics: Science and Systems*, 2014.

[4] S. Bao and S. Savarese. Semantic structure from motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2025–2032, 2011.

[5] M. Blaschko and C. Lampert. Object localization with global and local context kernels. In *British Machine Vision Conference (BMVC)*, pages 63.1–63.11, 2009.

[6] M. Cummins and P. Newman. Fab-map: Probabilistic localization and mapping in the space of appearance. *International Journal of Robotics Research (IJRR)*, 27 (6):647–665, 2008.

[7] C. Daganzo. *Multinomial probit: the theory and its application to demand forecasting*. Elsevier, 2014.

[8] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar. Next century challenges: Scalable coordination in sensor networks. In *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, pages 263–270. ACM, 1999.

[9] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J. Fernandez-Madrigal, and J. Gonzalez. Multi-hierarchical semantic maps for mobile robotics. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 2278–2283, 2005.

[10] D. Gálvez-López, M. Salas, J. Tardós, and J. Montiel. Real-time monocular object slam. *arXiv:1504.02398*, 2015.

[11] M. Huber. *Nonlinear Gaussian Filtering: Theory, Algorithms, and Applications*, volume 19. KIT Scientific Publishing, 2015.

[12] I. Ibragimov. On determining an infinitely divisible distribution function by its values on a half-line. *Theory of Probability & Its Applications*, 22(2):384–390, 1978.

[13] R. Ivanov, N. Atanasov, M. Pajic, I. Lee, and G. Pappas. Robust localization using context-aware filtering. In *Workshop on Multi-View Geometry in Robotics*, 2015.

[14] R. Ivanov, J. Weimer, A. Simpao, M. Rehman, and I. Lee. Early detection of critical pulmonary shunts in infants. In *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*, ICCPS '15, pages 110–119. ACM, 2015.

[15] S. Joshi and S. Boyd. Sensor selection via convex optimization. *Transactions on Signal Processing*, 57 (2):451–462, 2009.

[16] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82 (1):35–45, 1960.

[17] J. Levinson, M. Montemerlo, and S. Thrun. Map-based precision vehicle localization in urban environments. In *Robotics: Science and Systems*, volume 4, page 1. Citeseer, 2007.

[18] R. Mahler. *Statistical Multisource-Multitarget Information Fusion*. Artech House, Inc., 2007.

[19] B. Mortazavi, M. Pourhomayoun, H. Ghasemzadeh, R. Jafari, C. Roberts, and M. Sarrafzadeh. Context-aware data processing to enhance quality of measurements in wireless health systems: An application to MET calculation of exergaming actions. *Internet of Things Journal, IEEE*, 2(1):84–93, Feb 2015.

[20] H. Nickisch and C. Rasmussen. Approximations for binary gaussian process classification. *Journal of Machine Learning Research (JMLR)*, 9(Oct):2035–2078, 2008.

[21] A. Nüchter and J. Hertzberg. Towards semantic maps for mobile robots. *Robotics and Autonomous Systems*, 56(11):915–926, 2008.

[22] A. Pronobis. *Semantic Mapping with Mobile Robots*. dissertation, KTH Royal Institute of Technology, 2011.

[23] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[24] R. Salas-Moreno, R. Newcombe, H. Strasdat, P. Kelly, and A. Davison. Slam++: Simultaneous localisation and mapping at the level of objects. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1352–1359, 2013.

[25] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry. Kalman filtering with intermittent observations. *Automatic Control, IEEE Transactions on*, 49(9):1453–1464, 2004.

[26] H. Sorenson. *Kalman Filtering: Theory and Application*. IEEE Press, 1985.

[27] E. Sudderth, A. Ihler, W. Freeman, and A. Willsky. Nonparametric belief propagation. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 605–612, 2003.

[28] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT press, 2005.

[29] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin. On efficient sensor scheduling for linear dynamical systems. *Automatica*, 48(10):2482–2493, 2012.

[30] J. Wang, H. Zha, and R. Cipolla. Coarse-to-fine vision-based localization by indexing scale-invariant features. *IEEE Trans. on Systems, Man, and Cybernetics*, 36(2): 413–422, 2006.

[31] J. Williams. *Information Theoretic Sensor Management*. PhD thesis, MIT, 2007.




Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 15:23:46 UTC from IEEE Xplore. Restrictions apply.