

Towards Robust State Estimation by Boosting the Maximum Correntropy Criterion Kalman Filter with Adaptive Behaviors

Seyed Fakoorian^{*1,2}, Angel Santamaria-Navarro^{*2}, Brett T. Lopez^{*2},
Dan Simon¹ and Ali-akbar Agha-mohammadi²

Abstract— This work proposes a resilient and adaptive state estimation framework for robots operating in perceptually-degraded environments. The approach, called Adaptive Maximum Correntropy Criterion Kalman Filtering (AMCCKF), is inherently robust to corrupted measurements, such as those containing jumps or general non-Gaussian noise, and is able to modify filter parameters online to improve performance. Two separate methods are developed – the Variational Bayesian AM-CCKF (VB-AMCCKF) and Residual AMCCKF (R-AMCCKF) – that modify the process and measurement noise models in addition to the bandwidth of the kernel function used in MCCKF based on the quality of measurements received. The two approaches differ in computational complexity and overall performance which is experimentally analyzed. The method is demonstrated in real experiments on both aerial and ground robots and is part of the solution used by the COSTAR team participating at the DARPA Subterranean Challenge.

I. INTRODUCTION

Reliable operation of autonomous systems in diverse environments stresses the importance of accurate and robust state estimation. The phrase “robust estimation” is broad ([1]) but can be generally categorized as algorithms that can detect and mitigate the effects of measurement dropout, divergence, or discontinuities, i.e., jumps [2]. Several modifications to the well-known Kalman filter and its nonlinear counterparts – the extended Kalman filter (EKF) and unscented Kalman filter (UKF) – have been proposed. A recent method called maximum correntropy criterion Kalman filter (MCCKF), presented in [3], [4] and further developed in [5], [6], [7], [8], has proven to be effective at rejecting diverging and jumping measurements. Unlike the KF, which is derived using minimum mean square error (MMSE), the MCCKF uses the correntropy criterion [4] to capture the higher-order statistics of the measurement signal. The MCCKF will therefore outperform the KF when measurements contain non-Gaussian noise, i.e., the measurements are diverging or discontinuous. While the MCCKF is able to robustly handle non-Gaussian measurements, it relies on selecting filter parameters *a priori*, which can lead to suboptimal performance if not done carefully. This observation has led to works that modify – in an online fashion – some [9], [10] or all [11] filter parameters. While adapting all filter parameters is ideal,

an approach has yet to be developed that does not potentially amplify the effects of outlier measurements. To address these shortcomings, this work develops a framework combining MCCKF with online filter adaptation techniques to form a unified robust and adaptive estimation pipeline that can be easily integrated into existing autonomy pipelines.

Static noise models often employed in Kalman filtering is a common source of error in state estimation. This has led to significant interest in adaptive Kalman filters (AKFs) where the noise model is learned online with real measurement data. AKFs can be classified as either correlation, covariance matching, maximum likelihood, or Bayesian methods [12], [9]. The most common approach is the residual-based AKF (RAKF) [12] which is a maximum likelihood technique that estimates the noise covariance matrix over a sliding-window under the assumption the noise is Gaussian. More recent AKFs are based on Variational Bayesian techniques (VB) in which the noise covariance matrix is modeled as inverse Wishart distributions [13], [14]. To the best of the authors’ knowledge, an AKF method for non-Gaussian noise has yet to be developed.

Contribution: This paper presents a novel step towards robust state estimation using Kalman filters. The proposed filter, called adaptive-MCCKF (AMCCKF), combines the benefits of the AKF and MCCKF to achieve resilient and accurate odometry estimation of mobile robots. The proposed filter has the following features: a) online adaptation of the noise covariance matrices (similar to AKF); b) improved adaptation of the KB that *will not* amplify the effects of non-Gaussian noise (unlike [11]); c) produce reliable estimates when the measurement or process noise is not Gaussian, e.g., contains jumps. Two variants of the AMCCKF are derived: a VB-based AM-CCKF and a residual-based AM-CCKF. This fusion approach is part of the state estimation framework developed by team CoSTAR¹ for the DARPA Subterranean Challenge², Cave Circuit [15]. We show the validity of the proposed approach by fusing real data from experiments with a Clearpath Husky³ robot both running a loosely-coupled sensor fusion architecture.

The rest of this paper is organized as follows. Section II reviews the MCCKF formulations. Section III presents the VB-based AM-CCKF and residual-based AM-CCKF. Section IV describes the kinematic motion model and error state

* Authors contributed equally to this manuscript.

¹Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, OH, USA. {s.fakoorian; d.j.simon}@csuohio.edu.

²NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. {angel.santamaria.navarro; brett.t.lopez; aliakbar.aghamohammadi}@jpl.nasa.gov.

¹<https://costar.jpl.nasa.gov>

²<https://www.subtchallenge.com>

³<https://clearpathrobotics.com>

equations commonly used in mobile robot localization. Section V shows the experimental results. Section VI contains concluding remarks.

II. PRELIMINARIES

A standard assumption in Kalman filtering is the dynamics and measurements are corrupted by zero-mean Gaussian noise. If this assumption is violated then another method is required. The MCCKF uses higher-order statistics of the process and measurement noise to robustly eliminate the effects of non-Gaussian noise thereby preserving the Gaussian distribution of the states. Before presenting the main results of this work, the maximum correntropy criterion and its use in state fusion is briefly summarized.

Consider the nonlinear system

$$\begin{aligned}\mathbf{x}_k &= f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_{k-1}, \\ \mathbf{z}_k &= h(\mathbf{x}_k) + \mathbf{v}_k,\end{aligned}\quad (1)$$

where $\mathbf{x}_k \in \mathbb{R}^n$, $\mathbf{u}_k \in \mathbb{R}^p$, and $\mathbf{z}_k \in \mathbb{R}^m$ are the state vector, propagation inputs and observations, respectively. The nonlinear dynamics and measurement model are $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, respectively. The process and measurement noise are \mathbf{w}_k and \mathbf{v}_k and generally treated as zero-mean independent white noise processes with covariance matrices \mathbf{Q}_k and \mathbf{R}_k , respectively. The MCCKF equations to estimate the system in (1) are shown below (see [10] for a more detailed derivation).

System model propagation (prior estimation):

$$\hat{\mathbf{x}}_{k|k-1} = f(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_k), \quad (2)$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^\top + \mathbf{Q}_k. \quad (3)$$

Observation correction (posterior estimation):

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \tilde{\mathbf{y}}_k, \quad (4)$$

$$\mathbf{K}_k = \left(\mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_k^\top \mathbf{C}_k \mathbf{R}_k^{-1} \mathbf{H}_k \right)^{-1} \mathbf{H}_k^\top \mathbf{C}_k \mathbf{R}_k^{-1}, \quad (5)$$

$$\begin{aligned}\mathbf{C}_k = \text{diag} \left[\mathbf{G}_{\sigma,k} \left(\|\tilde{\mathbf{y}}_{1,k}\|_{\mathbf{R}_{1,k}^{-1}} \right), \dots, \mathbf{G}_{\sigma,k} \left(\|\tilde{\mathbf{y}}_{m,k}\|_{\mathbf{R}_{m,k}^{-1}} \right) \right], \quad (6)\end{aligned}$$

$$\begin{aligned}\mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^\top \\ &\quad + \mathbf{K}_k \mathbf{R}_k (\mathbf{K}_k)^\top,\end{aligned}\quad (7)$$

where $\tilde{\mathbf{y}}_k = \mathbf{z}_k - h(\hat{\mathbf{x}}_{k|k-1})$ and $\tilde{\mathbf{y}}_{j,k} = \mathbf{z}_{j,k} - h_j(\hat{\mathbf{x}}_{k|k-1})$, with $j \in \{1, \dots, m\}$; $\mathbf{G}_{\sigma,k}(\cdot) = \exp\left(\frac{-\|\cdot\|^2}{2\sigma_k^2}\right)$, with σ_k the size of the KB; $\mathbf{K}_k \in \mathbb{R}^{n \times m}$ is the MCCKF gain; $\mathbf{C}_k \in \mathbb{R}^{m \times m}$ is the correntropy gain; $\hat{\mathbf{x}}_{k|k-1}$ is the *a priori* estimate of the state $\hat{\mathbf{x}}_{k|k}$, which integrates measurements up to (and including) time $k-1$ and has covariance $\mathbf{P}_{k|k-1}$. The *a posteriori* estimate of the state $\hat{\mathbf{x}}_{k|k-1}$ is $\hat{\mathbf{x}}_{k|k}$, which is based on measurements up to (and including) time k and has covariance $\mathbf{P}_{k|k}$. Note that $\|\cdot\|_A$ denotes the A -weighted two-norm of a vector. The j -th row of the vector resulting from h is h_j and, similarly, $\mathbf{R}_{j,k}$ is the j -th element on the diagonal of \mathbf{R}_k .

It can be seen that when the j -th measurement is disturbed by a large outlier, the j -th element of the correntropy gain \mathbf{C}_k in (6) goes to zero and prevents estimator divergence. Further, the innovation term of the j -th measurement can also be shown to be zero in this scenario. In Section III we propose two different methods for adapting the covariance noise matrices $\hat{\mathbf{R}}_k$ and $\hat{\mathbf{Q}}_k$. We then find an adaptive method to select the KB σ_k . For the sake of simplicity, in the following we keep only the symbol $\hat{\cdot}$ for the main variables to estimate, i.e., $\hat{\mathbf{x}}$, $\hat{\sigma}$, $\hat{\mathbf{R}}$ and $\hat{\mathbf{Q}}$.

Two comments are in order. First, note that other cost functions besides the MCC can be incorporated into the KF to reject outliers. For instance, Huber, Hampel or IGG (Institute of Geodesy and Geophysics) [16]. However, as discussed in [7] and [16], MCC-based filters will outperform these filters especially when the underlying system is disrupted by impulsive noise. Fundamentally, this is a result of large outliers being completely removed by the MCC as opposed to only being truncated by other loss functions. And second, the MCCKF is susceptible to the same effects of linearization error and unmodeled dynamics as the EKF when outliers are not present. However, since linearization error and excitation of unmodeled dynamics are exacerbated with measurement outliers, the MCCKF will outperform traditional EKF algorithms since the MCCKF is able to directly compensate for non-Gaussian noise. Additional robustness and performance improvement are also possible by adapting the process and measurement noise covariance matrices. These points will be supported by the experiments presented in Section V.

III. PROPOSED METHOD

A. Variational Bayesian based AMCCKF

A sliding-window variational adaptive Kalman filter (SWVAKF) [14] scheme includes two steps: the forward KF (identical with the standard KF); and the backward Kalman smoother (KS). The forward KF calculates state estimates based on previous estimates of noise covariance matrices. The backward KS provides an approximate smoothing posterior probability density function (PDF) over the sliding-window state vectors. This PDF is later combined with the posterior PDF of noise covariance matrices to make a joint posterior PDF of $p(\hat{\mathbf{x}}_k, \hat{\mathbf{R}}_k, \hat{\mathbf{Q}}_k | \mathbf{z}_\omega)$. Solving this joint PDF results in an online estimate of \mathbf{R}_k and \mathbf{Q}_k . In SWVAKF, the variational Bayesian (VB) method is used to estimate online the joint posterior PDF

$$p(\hat{\mathbf{x}}_k, \hat{\mathbf{R}}_k, \hat{\mathbf{Q}}_k | \mathbf{z}_\omega) \approx q(\hat{\mathbf{x}}_k | \mathbf{z}_\omega) q(\hat{\mathbf{R}}_k | \mathbf{z}_\omega) q(\hat{\mathbf{Q}}_k | \mathbf{z}_\omega), \quad (8)$$

where $q(\cdot)$ is an approximation of $p(\cdot)$ and \mathbf{z}_ω is a notation simplification of $\mathbf{z}_{k-\omega:k}$. Since both $\hat{\mathbf{R}}_k$ and $\hat{\mathbf{Q}}_k$ are the covariance matrices representing Gaussian PDFs with zero mean, their posterior distributions can be modeled as inverse-Wishart PDFs (IW) to guarantee conjugate inference

$$q(\hat{\mathbf{R}}_k) = \text{IW}(\hat{\mathbf{R}}_k | \hat{b}_{k|k}, \hat{\mathbf{B}}_{k|k}), \quad (9)$$

$$q(\hat{\mathbf{Q}}_k) = \text{IW}(\hat{\mathbf{Q}}_k | \hat{t}_{k|k}, \hat{\mathbf{T}}_{k|k}), \quad (10)$$

where $\hat{t}_{k|k}$ and $\hat{b}_{k|k}$ are the parameters related to the *degrees-of-freedom* (DoFs); and $\hat{\mathbf{T}}_{k|k}$ and $\hat{\mathbf{B}}_{k|k}$ are the inverse scale matrices, all described by

$$\hat{t}_{k|k} = \rho \hat{t}_{k|k-1} + \varpi, \quad \hat{\mathbf{T}}_{k|k} = \rho \hat{\mathbf{T}}_{k|k-1} + \sum_{j=k-\varpi}^k \mathbf{O}_j, \quad (11)$$

$$\hat{b}_{k|k} = \rho \hat{b}_{k|k-1} + \varpi, \quad \hat{\mathbf{B}}_{k|k} = \rho \hat{\mathbf{B}}_{k|k-1} + \sum_{j=k-\varpi}^k \mathbf{M}_j, \quad (12)$$

where ρ is the forgetting factor and recommended to be in the interval $\rho \in [0.9, 1]$. The matrices \mathbf{O}_j and \mathbf{M}_j are

$$\mathbf{O}_j = \mathbf{P}_{j|k} - \mathbf{F}_j \mathbf{P}_{j-1,j|k} - \mathbf{P}_{j-1,j|k}^\top \mathbf{F}_j^\top + \mathbf{F}_j \mathbf{P}_{j-1,j|k} \mathbf{F}_j^\top + \tilde{\mathbf{x}}_j | k \tilde{\mathbf{x}}_j^\top | k, \quad (13)$$

$$\mathbf{M}_j = \mathbf{r}_j \mathbf{r}_j^\top + \mathbf{H}_j \mathbf{P}_{j|k} \mathbf{H}_j^\top. \quad (14)$$

Estimates $\hat{\mathbf{R}}_k$ and $\hat{\mathbf{Q}}_k$ can then be obtained ([14])

$$\hat{\mathbf{R}}_k = \hat{\mathbf{B}}_{k|k} / \hat{b}_{k|k}, \quad (15)$$

$$\hat{\mathbf{Q}}_k = \hat{\mathbf{T}}_{k|k} / \hat{t}_{k|k}. \quad (16)$$

The SWVAKF can be combined with the MCCKF to improve its robustness to outliers. From (8), $q(\hat{\mathbf{x}}_k)$ can be approximated as a normal Gaussian distribution, i.e., $q(\hat{\mathbf{x}}_k) = \mathcal{N}(\hat{\mathbf{x}}_k | \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$, where the mean vector $\hat{\mathbf{x}}_{k|k}$ can be obtained under the MCC principles [4]. Thus, we can use (2)–(7) instead of the forward KF step in the SWVAKF (see [14] for further details on SWVAKF design). Since the SWVAKF was derived based on the assumption that the noise has a Gaussian distribution, we propose to use the correntropy gain in (14) to reduce the effect of corrupted measurements while estimating \mathbf{R}_k , e.g., when the noise is non-Gaussian. Defining the *unweighted* correntropy gain $\mathbf{L}_k := \text{diag}[\mathbf{G}_{\sigma,k}(\|\tilde{\mathbf{y}}_{1,k}\|), \dots, \mathbf{G}_{\sigma,k}(\|\tilde{\mathbf{y}}_{m,k}\|)]$, we can approximate the auxiliary matrix \mathbf{M} (14) by right and left multiplying the unweighted correntropy gain \mathbf{L}_k with the first term leading to

$$\mathbf{M}_j = \mathbf{L}_j \mathbf{r}_j \mathbf{r}_j^\top \mathbf{L}_j + \mathbf{H}_j \mathbf{P}_{j|k} \mathbf{H}_j^\top. \quad (17)$$

Similar to the above-derived residual-based AMCCKF, with $\mathbf{L}_j \approx \mathbf{I}$ (e.g., in the presence of Gaussian noise) (17) reduces to (14). However, with non-Gaussian noise $\mathbf{L}_k \approx \mathbf{0}$ which reduces the impact of corrupted measurements in the update of $\hat{\mathbf{R}}_k$. Note that the backward KF step in the VB-based AMCCKF is the same as in the traditional SWVAKF. Algorithm 1 shows the modified VB-based AMCCKF design.

Before concluding, it is instructive to briefly discuss the computational complexity of the SWVAKF and VB-AMCCKF. The computation complexity of both SWVAKF and VB-AMCCKF is highly dependant on the window size in the KS step. The computation grows linearly as the ϖ gets larger [14]. It is suggested to choose ϖ within 5 to 20 range for optimal computation complexity [14].

Algorithm 1 Variational Bayesian-based AMCCKF

Initialization:

1: $\hat{\mathbf{x}}_0, \mathbf{P}_0, \hat{\mathbf{R}}_0, \hat{\mathbf{Q}}_0, t_0, \mathbf{T}_0, b_0$, and \mathbf{B}_0

Forward MCCKF: Using (2)–(7)

2: $[\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}] = MCKF(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \hat{\mathbf{Q}}_{k-1}, \hat{\mathbf{R}}_{k-1}, \mathbf{F}_k, \mathbf{H}_k, \mathbf{z}_k)$

Backward KS:

3: **for** $j = [k-\varpi : k]$ **do**
 4: $\mathbf{P}_{j|j-1} = \mathbf{F}_k \mathbf{P}_{j-1|j-1} \mathbf{F}_k^\top + \hat{\mathbf{Q}}_j$
 5: $\mathbf{G}_{j-1} = \mathbf{P}_{j-1|j-1} \mathbf{F}_k^\top (\mathbf{P}_{j|j-1})^{-1}$
 6: $\hat{\mathbf{x}}_{j-1|k} = \hat{\mathbf{x}}_{j-1|j-1} + \mathbf{G}_{j-1} (\hat{\mathbf{x}}_{j|k} - \mathbf{F}_k \hat{\mathbf{x}}_{j|j-1})$
 7: Obtain \mathbf{O}_j and \mathbf{M}_j using (13) and (17)

8: Obtain $\mathbf{M}_{k-\varpi}$ using (17)

Estimation of \mathbf{Q}_k and \mathbf{R}_k :

9: Obtain the Wishart parameters using (11)–(12)
 10: Update $\hat{\mathbf{R}}_k$ and $\hat{\mathbf{Q}}_k$ using (15)–(16)

B. Residual-based AMCCKF

The SWVAKF presented in Section III-A can be computationally expensive due to the backward KS step. This section presents a residual-based approach that has lower computational complexity with similar accuracy making it more suitable for mobile robots with limited onboard computation. The traditional residual-based adaptive KF (RAKF) estimates \mathbf{R}_k and \mathbf{Q}_k assume the innovation sequence of the Kalman filter is a white process [13]. The RAKF then solves a maximum likelihood (ML) problem [12]. We propose to modify the ML problem by incorporating the correntropy gain directly in the ML optimization. In this case, the probability density function of the measurements \mathbf{z} with the adaptive parameter α at the epoch k is

$$p(\mathbf{z} | \alpha) = \frac{1}{\sqrt{(2\pi)^m |\Gamma_k|}} e^{\frac{-1}{2} \mathbf{L}_k \mathbf{y}_k \Gamma_k^{-1} \mathbf{y}_k^\top \mathbf{L}_k}, \quad (18)$$

where m is the number of measurements; \mathbf{L}_k is the correntropy gain (6), without the weighting matrix \mathbf{R}_k^{-1} ; and Γ_k is the covariance matrix of the innovation sequence, which is dependent on the adaptive parameter α . Since $\Gamma_k := \mathbb{E}[\mathbf{y}_k \mathbf{y}_k^\top]$ then direct substitution gives [17]

$$\Gamma_k = \mathbf{H} \mathbf{P}_{k|k-1} \mathbf{H}^\top + \hat{\mathbf{R}}_k. \quad (19)$$

The main principle behind an adaptive Kalman filter is to determine Γ_k and its partial derivative with respect to α . One can then formulate an online stochastic natural gradient descent on the log-likelihood of the observations. Moreover, we are interested in averaging the adaptation of $\hat{\mathbf{R}}_k$ and $\hat{\mathbf{Q}}_k$ over a sliding window of length ϖ . First, we take the natural logarithm of (18) so

$$\mathbf{J} = \sum_{j=k-\varpi}^k \ln |\Gamma_j| + \sum_{j=k-\varpi}^k \mathbf{L}_j \mathbf{y}_j \Gamma_j^{-1} \mathbf{y}_j^\top \mathbf{L}_j. \quad (20)$$

Then, maximizing (18) translates into minimizing (20) with respect to α . Therefore, one can compute the following gradient

$$\frac{\partial \mathbf{J}}{\partial \boldsymbol{\alpha}_k} = \sum_{j=j_0}^k \left[\text{tr} \left(\boldsymbol{\Gamma}_j^{-1} \frac{\partial \boldsymbol{\Gamma}_j}{\partial \boldsymbol{\alpha}_k} \right) - \mathbf{L}_j \mathbf{y}_j \boldsymbol{\Gamma}_j^{-1} \frac{\partial \boldsymbol{\Gamma}_j}{\partial \boldsymbol{\alpha}_k} \boldsymbol{\Gamma}_j^{-1} \mathbf{y}_j^\top \mathbf{L}_j \right]. \quad (21)$$

Substituting (19) and its derivative (see [12]) into (21), the ML equation for the AMCCKF results in

$$\sum_{j=k-\varpi}^k \text{tr} \left([\boldsymbol{\Gamma}_j^{-1} - \mathbf{L}_j \boldsymbol{\Gamma}_j^{-1} \mathbf{y}_j \mathbf{y}_j^\top \boldsymbol{\Gamma}_j^{-1} \mathbf{L}_j] \left[\frac{\partial \hat{\mathbf{R}}_j}{\partial \boldsymbol{\alpha}_k} \right. \right. \\ \left. \left. + \mathbf{H}_j \frac{\partial \hat{\mathbf{Q}}_j}{\partial \boldsymbol{\alpha}_k} \mathbf{H}_j^\top \right] \right) = 0. \quad (22)$$

To obtain an explicit equation for $\hat{\mathbf{R}}_k$, one first assumes that $\hat{\mathbf{Q}}_k$ is known, and that $\hat{\mathbf{R}}_k$ has values only along its main, i.e., $\hat{\mathbf{R}}_{i,k} = \alpha_i$. Then, (22) reduces to

$$\sum_{j=k-\varpi}^k \text{tr} (\boldsymbol{\Gamma}_j^{-1} [\boldsymbol{\Gamma}_j - \mathbf{L}_j \mathbf{y}_j \mathbf{y}_j^\top \mathbf{L}_j] \boldsymbol{\Gamma}_j^{-1}) = 0. \quad (23)$$

The solution to (23) can be obtained under a sliding-window (average) as

$$\boldsymbol{\Gamma}_k = \frac{1}{\varpi} \sum_{j=k-\varpi}^k \mathbf{L}_j \mathbf{y}_j \mathbf{y}_j^\top \mathbf{L}_j. \quad (24)$$

Hence, the adaptive innovation-based expression for $\hat{\mathbf{R}}_k$ can be achieved by substituting (24) into (19), resulting in

$$\hat{\mathbf{R}}_k = \boldsymbol{\Gamma}_k - \mathbf{H} \mathbf{P}_{k|k-1} \mathbf{H}^\top. \quad (25)$$

To guarantee that $\hat{\mathbf{R}}_k$ remains positive-definite, we can adapt (24) and (25) to use the residual (i.e., residual-based adaptive method) $\mathbf{r}_k = \mathbf{z}_k - h(\hat{\mathbf{x}}_{k|k})$, obtaining

$$\boldsymbol{\Gamma}_k = \frac{1}{\varpi} \sum_{j=k-\varpi}^k \mathbf{L}_j \mathbf{r}_j \mathbf{r}_j^\top \mathbf{L}_j, \quad (26)$$

$$\hat{\mathbf{R}}_k = \boldsymbol{\Gamma}_k + \mathbf{H} \mathbf{P}_{k|k} \mathbf{H}^\top. \quad (27)$$

The reader is referred to [12] for more details on the derivation of (26)-(27).

If non-Gaussian noise is present within the sliding-window, for example in one of the measurements, then the residual is no longer normal leading to an accumulative error in traditional RAKF. As a result, the estimated $\hat{\mathbf{R}}_k$ for that particular window of measurements is not accurate even though the noise for the most recent measurement might be of Gaussian distribution. The usage of the correntropy gain \mathbf{L}_k mitigates the effects preventing the divergence of the states and $\hat{\mathbf{R}}_k$. We should also note that \mathbf{L}_k is approximately the identity matrix when the noise is Gaussian ($\mathbf{L}_k \approx \mathbf{I}$), thus the equation for adapting $\hat{\mathbf{R}}_k$ is reduced to a traditional RAKF.

To compute an adaptive $\hat{\mathbf{Q}}_k$ we can follow a similar strategy as for $\hat{\mathbf{R}}_k$, but this time assuming $\hat{\mathbf{R}}_k$ is known

and $\hat{\mathbf{Q}}_{i,k} = \alpha_i$ in (22). In this case, the adaptive equation for \mathbf{Q}_k can be approximated as $\hat{\mathbf{Q}}_k = \mathbf{K}_k \boldsymbol{\Gamma}_k \mathbf{K}_k^\top$, with $\boldsymbol{\Gamma}_k$ given in (24). The developments to obtain $\hat{\mathbf{Q}}_k$ are similar but omitted for brevity (see e.g., [12]).

C. Adaptive Kernel Bandwidth selection:

An important parameter in MCCKF often selected through trial and error with real sensor data is the kernel bandwidth (KB). A large KB limits the ability of the filter to reject outliers while a small KB can cause slow convergence or, in some cases, divergence of the filter [4]. This work proposes to *adapt* the KB online based on the quality of measurements received. Drawing inspiration from [10] for the dynamic selection of the KB size and motivated by the adaptation of $\hat{\mathbf{R}}_k$ and $\hat{\mathbf{Q}}_k$, we define

$$\hat{\sigma}_{\mu,k} = \left(\|\mathbf{y}_{\mu,k} \mathbf{y}_{\mu,k}^\top\|_{\hat{\mathbf{R}}_{\mu,k-1}} + \mathbf{H}_\mu \mathbf{P}_{k|k-1} \mathbf{H}_\mu^\top \right)^{-1}, \quad (28)$$

where μ is the index of the particular dimension of the observation being evaluated (e.g., if we observe only 3D position, $\mu \in \{1, 2, 3\}$); and $\mathbf{P}_{k|k-1}$ is obtained from (3), using $\hat{\mathbf{Q}}_k$. Intuitively, when a measurement is not corrupted non-Gaussian noise then $\hat{\sigma}_{\mu,k}$ in (28) becomes large and $\mathbf{L}_{\mu,k}, \mathbf{C}_{\mu,k} \approx \mathbf{I}$. However, if a measurement classified as an large outlier, $\hat{\sigma}_{\mu,k}$ will become small and $\mathbf{L}_{\mu,k}, \mathbf{C}_{\mu,k} \approx \mathbf{0}$ preventing a catastrophic state correction. Moreover, an adaptive KB has the advantage of measurement-specific adaptation, i.e., independent for each observation dimension.

Notice how the adaptive KB in (28) is the sum of two positive terms, therefore $\hat{\sigma}_{\mu,k}$ is always positive and well-defined. This is not the case for [11], where they proposed a KB adaptation that can reach negative values thus creating an inconsistency that the MCCKF cannot handle. KB adaption can be used in both VB- and R-based AMCCKF.

IV. ROBOT STATE ESTIMATION ON SE(3)

A critical capability for autonomous mobile robots is robust ego-motion estimation. To this end, we now show how the presented AMCCKF schemes can be used to robustly localize a mobile platform. Existing filter-based localization approaches can be classified as loosely- or tightly-coupled. The main difference being the type of measurements used in the filter. This work takes a loosely-coupled approach where the proposed AMCCKF algorithm is the main fusion engine that use an Inertial Measurement Unit (IMU) to propagate the nonlinear model and 6DoFs odometry estimates are used to correct this propagation (i.e., odometry estimations computed by one or several external front-end modules). In both propagation and correction steps we use the estimates $\hat{\mathbf{Q}}$, $\hat{\mathbf{R}}$ and $\hat{\sigma}_\mu$ described in the previous section.

A. System kinematics in discrete time

State and covariance propagation requires a discretized version of the kinematic motion model of the robot to infer its true-state \mathbf{x} . However, it is common to estimate the error-state instead of the true-state, where the error-state $\delta\hat{\mathbf{x}}$ values represent the discrepancy between the nominal-state $\hat{\mathbf{x}}$ and the true-state values \mathbf{x} [18]. This type of propagation and

update model is called Error-State Kalman Filter (ESKF) or, less frequently, Indirect Kalman Filter. In the following developments we integrate the principles of ESKF within the AMCCKF design. The advantages of using the error-state kinematic model have been discussed in [19].

First, we define the nominal state kinematics in discrete-time with a first-order (Euler) integration (i.e., a Taylor series expansion, truncating the series at first grade) as

$$\hat{\mathbf{p}}_k = \hat{\mathbf{p}}_{k-1} + \hat{\mathbf{v}}_{k-1} \Delta t, \quad (29)$$

$$\hat{\mathbf{v}}_k = \hat{\mathbf{v}}_{k-1} + [\hat{\mathcal{R}}_{k-1}(\mathbf{a}_r - \mathbf{a}_b) + \mathbf{g}] \Delta t, \quad (30)$$

$$\hat{\mathbf{q}}_k = \hat{\mathbf{q}}_{k-1} \otimes \mathbf{q}\{(\boldsymbol{\omega}_r - \boldsymbol{\omega}_b)\Delta t\}, \quad (31)$$

where \mathbf{p} and \mathbf{q} are the robot's 3D position and orientation (expressed as Hamiltonian quaternion with convention $[q_w, q_x, q_y, q_z]$), respectively. The operation $\mathbf{q}\{i\}$ represents the quaternion created from the 3D rotation vector i and \mathcal{R} represents a 3D rotation matrix. The 3D linear velocity vector is \mathbf{v} . The raw accelerometer and gyroscope measurements are \mathbf{a}_r and $\boldsymbol{\omega}_r$ with biases \mathbf{a}_b and $\boldsymbol{\omega}_b$. Note that the biases are not directly estimated in this work as the IMU used in the experiments estimates them internally. Thus the corrected measurements produced by the IMU are denoted as $\mathbf{a}_s := \mathbf{a}_r - \mathbf{a}_b$ and $\boldsymbol{\omega}_s := \boldsymbol{\omega}_r - \boldsymbol{\omega}_b$. Finally, Δt is the discretization time step (time differential).

The error state dynamics of the discrete-time kinematic model are given by

$$\delta\hat{\mathbf{p}}_k = \delta\hat{\mathbf{p}}_{k-1} + \delta\hat{\mathbf{v}}_{k-1} \Delta t, \quad (32)$$

$$\delta\hat{\mathbf{v}}_k = \delta\hat{\mathbf{v}}_{k-1} - \left(\left[\hat{\mathcal{R}}_{k-1} \mathbf{a}_s \right]_\times \delta\hat{\theta}_{k-1} \right) \Delta t, \quad (33)$$

$$\delta\hat{\theta}_k = \delta\hat{\theta}_{k-1}, \quad (34)$$

where we use a minimal orientation error $\delta\theta \in \mathfrak{so}(3) \subset \mathbb{R}^3$ living in the tangent space tangent of $SO(3)$ manifold (i.e., in its Lie algebra $\mathfrak{so}(3)$).

B. AMCCKF propagation and correction

The discrete-time kinematic model in (29)–(31) is propagated at the arrival of every new IMU measurement. Since the dynamics are nonlinear, the dynamics are linearized to obtain the Jacobian used to propagate the state covariance matrix. Note that in ESKF filters the prior estimate of the error-state is considered zero (i.e., $\delta\hat{\mathbf{x}}_{k|k-1} = 0$) because we assume the error mean $\delta\hat{\mathbf{x}}_{k|k-1}$ starting and remaining at zero until a new correction observation is received. Notice how the evaluation of (3) also includes \mathbf{Q}_k , which in our case corresponds to the estimated $\hat{\mathbf{Q}}_k$ through either the VB- or residual-based AMCCKF methods described in the previous section.

For the correction step, the filter state is updated at every arrival of an external odometry estimate. Hence, the observation model corresponds to a direct update to the robot position, orientation and linear velocity. The linearized observation model \mathbf{H}_k is required to update the error state and to compute the adaptive covariance of the measurement noise $\hat{\mathbf{R}}_k$. Computing the Jacobian of the measurement model, we



Fig. 1. Clearpath Husky robot is equipped with three RealSense T265 camera trackers (front, left, and right) and Vectornav VN100 inertial measurements units.

can evaluate the AMCCKF correction equations from (4)–(7) with the respective adaptive matrices $\hat{\mathbf{R}}_k$. Note that while we adapt only one $\hat{\mathbf{Q}}_k$ for the propagation step, we have to dynamically compute one $\hat{\mathbf{R}}_k$ for each observation source. For example, if we are accepting corrections from three external odometry estimation sources, we need to adapt three different $\hat{\mathbf{R}}_k$. We should note that the odometry measurement sensors are independent and AMCCKF is able to fuse them based on the arrival of each sensor.

A final step of the filter after every correction is to update the nominal-state $\hat{\mathbf{x}}_k$ with the observed error state $\delta\hat{\mathbf{x}}_k$ using the composition $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} \oplus \delta\hat{\mathbf{x}}_{k|k}$, where the operator \oplus stands for a generic addition operation to the different state components. Since $\delta\hat{\mathbf{x}}_k$ is reset to zero at the beginning of each new filter iteration, one can obtain the estimate of the true state as $\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_k$.

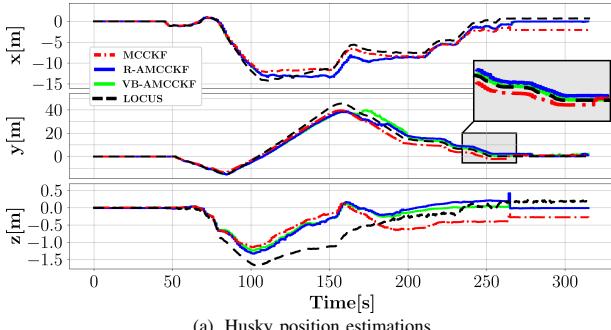
V. EXPERIMENTAL VALIDATION

The proposed method was evaluated with real experiments with a Clearpath Husky rover, able to traverse from indoor to outdoor scenarios and navigate large distances. The robot, shown in Fig. 1, is part of the solution defined by the team COSTAR for the Cave Circuit in the DARPA Subterranean Challenge. Our goal is to produce a smooth state estimate by fusing multiple odometry sources in a loosely-coupled fashion. Commercial-of-the-shelf visual-inertial odometry (VIO) from RealSense – the T265 camera⁴ – were mounted to the ground robot which can be seen in Fig. 1. The Husky had three T265 cameras; no cameras had overlapping field-of-view so they were treated as independent measurements. Moreover, the robot is equipped with a Vectornav VN100 IMU sensor⁵. The RealSense T265 and Vectonav IMU operate at 200Hz. Note that the gyro and accelerometer biases are not estimated since the VN100 IMU used in the experiments compensates for them internally.

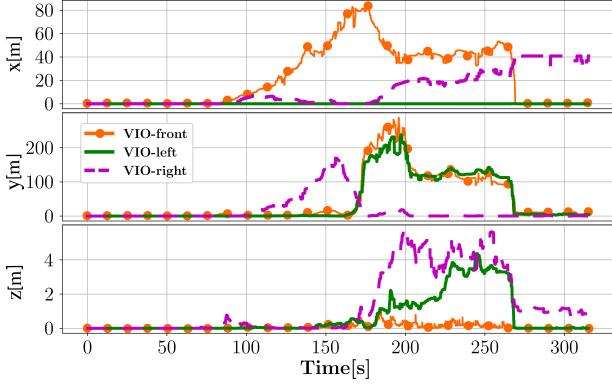
In the following experiments, the kinematic model was propagated with the IMU and corrected with each VIO source. The noise covariance matrix for each camera was initialized to $\hat{\mathbf{R}}_0 = 0.01 \times \mathbf{I}_9$. In addition, a window size of $\varpi = 10$ was chosen for both the VB-AMCCKF and R-AMCCKF. Other parameters for the VB-AMCCKF were set

⁴<https://dev.intelrealsense.com/docs/tracking-camera-t265-datasheet>

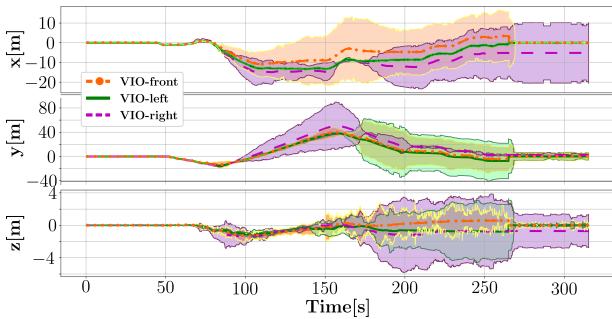
⁵<https://www.vectonav.com>



(a). Husky position estimations.



(b). Temporal variation of the adaptive kernel bandwidth (KB) inverse.



(c). The VIO estimations with their 3-sigma bands from $\hat{R}_{i,k}$.

Fig. 2. The VIO sensors are fused by VB-AMCCKF and R-AMCCKF. The position of the ground robot is shown in (a), ± 3 times of the square root of estimated covariance of each VIO sensors is shown in (b), and the inverse of the adaptive KB is shown in (c). Note that we only show the position of the robot here for the purpose of illustration, however same results hold for rest of the states.

to $\rho = 0.97$, $\hat{t}_{0|0} = \hat{b}_{0|0} = 0$, $\hat{T}_{0|0} = \mathbf{0}_9$, and $\hat{B}_{0|0} = \mathbf{0}_9$. Even though both platforms carry other sensors, only the 3D LiDAR (Velodyne VLP-16⁶) mounted on the Husky was used for comparison purposes. The following experiments were carried out at the NASA Jet Propulsion Laboratory (JPL).

Both the VB-AMCCKF and R-AMCCKF methods were tested on the Husky robot during an experiment where the robot executed a several-meter traverse in an urban environment. In this experiment the results are compared to a LiDAR-based odometry estimation algorithm called LOCUS [20]. The robot was driven over non-flat terrain to intentionally degrade the performance of the T265 cameras.

Fig. 2a first compares the traditional MCCKF (manually tuned with a fixed kernel size and covariance matrices) and the proposed VB-AMCCKF and R-AMCCKF with adapt-

⁶<https://velodynelidar.com/products/puck>

TABLE I: CPU Usage Comparison.

		Min%	Max%	Avg%	Std%
Husky	R-AMCCKF	32	74	46	14
	VB-AMCCKF	39	86	65	16

tive KB; the noise covariance matrices were kept static in this test. Here, the traditional MCCKF cannot reach the accuracy of the VB-AMCCKF or R-AMCCKF because the MCCKF lacks adaptive behavior; both VB-AMCCKF and R-AMCCKF generate similar estimates to LOCUS at the end of the trajectory. The inverse of the adaptive KB (Fig. 2b) shows the temporal weighting of the sensors in the filter. Smaller values correspond to larger a correntropy gain (i.e., $C_k \rightarrow I$ and more impact in the estimation) and vice versa. Hence, the inverse can be interpreted as a metric for anticipated sensor health; a smaller value corresponds to a sensor that is unlikely to produce an unhealthy measurement, i.e., a healthy sensor. VB-AMCCKF has shown better precision than R-AMCCKF but requiring more computation. We compare the average percentage of the CPU usage between the VB-AMCCKF and R-AMCCKF in Table I. As expected the VB-AMCCKF requires more CPU effort.

As previously stated, one of the advantages of the AMCCKF scheme is the adaptation of the observation noise covariances \hat{R} . Fig. 2c shows how the R-AMCCKF adapts for each VIO sensor. The modification of individual dimensions on the covariance matrices allow the AMCCKF to take advantage of the “best” available estimate for each individual dimension. For instance, notice how VIO-left performs best in the estimation for x position, while a combination of front and right cameras prevail for y , and VIO-front for z . Fig. 2 further demonstrated the effectiveness of the developed VB-AMCCKF and R-AMCCKF.

VI. CONCLUSION

This paper proposes two AMCCKF algorithms, based on the Variational Bayesian (VB-AMCCKF) and Residual (R-AMCCKF) techniques, as a step toward robust sensor fusion. The two designs can be selected based on desired accuracy and available onboard computation. Fundamentally, AMCCKF leverage the advantages of MCCKF filters, which show improved robustness with respect to traditional Kalman filters, and extended the design to incorporate adaptive behaviors. The adaptations include a dynamic selection of a) the kernel bandwidth size; b) the system noise covariance matrix; and c) the measurement noise covariance matrix. The addition of online adaptation reduce, the effects of rigid (and usually) inaccurate user-defined parameters. The validity of the proposed filters was demonstrated through real experiments on a Husky ground robot.

ACKNOWLEDGEMENTS

This research work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Copyright 2020 California Institute of Technology. U.S. Government sponsorship acknowledged.

REFERENCES

- [1] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, 2012.
- [2] A. Santamaria-navarro, R. Thakker, D. D. Fan, B. Morrell, and A. Agha-mohammadi, "Towards resilient autonomous navigation of drones," in *International Symposium on Robotics Research*, 2019.
- [3] R. Izanloo, S. A. Fakoorian, H. S. Yazdi, and D. Simon, "Kalman filtering based on the maximum correntropy criterion in the presence of non-gaussian noise," in *IEEE Annual Conference on Information Science and Systems*, 2016, pp. 500–505.
- [4] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, 2017.
- [5] M. V. Kulikova, "Square-root algorithms for maximum correntropy estimation of linear discrete-time systems in presence of non-gaussian noise," *Systems & Control Letters*, vol. 108, pp. 8–15, 2017.
- [6] Y. Huang, Y. Zhang, N. Li, Z. Wu, and J. A. Chambers, "A novel robust student's t-based Kalman filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1545–1554, 2017.
- [7] G. Wang, N. Li, and Y. Zhang, "Maximum correntropy unscented Kalman and information filters for non-Gaussian measurement noise," *Journal of the Franklin Institute*, vol. 354, no. 18, pp. 8659–8677, 2017.
- [8] M. V. Kulikova, "Chandrasekhar-based maximum correntropy Kalman filtering with the adaptive kernel size selection," *IEEE Transactions on Automatic Control*, vol. 65, no. 2, pp. 741–748, 2019.
- [9] S. Akhlaghi, N. Zhou, and Z. Huang, "Adaptive adjustment of noise covariance in Kalman filter for dynamic state estimation," in *IEEE Power Energy Society General Meeting*, 2017, pp. 1–5.
- [10] S. Fakoorian, R. Izanloo, A. Shamshirgaran, and D. Simon, "Maximum correntropy criterion Kalman filter with adaptive kernel size," in *2019 IEEE National Aerospace and Electronics Conference*, 2019, pp. 581–584.
- [11] S. Li, B. Xu, L. Wang, and A. A. Razzaqi, "Improved maximum correntropy cubature Kalman filter for cooperative localization," *IEEE Sensors Journal*, pp. 1–1, 2020.
- [12] A. Mohamed and K. Schwarz, "Adaptive Kalman filtering for INS/GPS," *Journal of geodesy*, vol. 73, no. 4, pp. 193–203, 1999.
- [13] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Transactions on Automatic Control*, vol. 63, no. 2, pp. 594–601, 2017.
- [14] Y. Huang, F. Zhu, G. Jia, and Y. Zhang, "A slide window variational adaptive Kalman filter," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2020.
- [15] A. Agha-mohammadi, K. Otsu, B. Morrell, and et. al., "NeBula: Quest for robotic autonomy in challenging environments; TEAM CoSTAR at the DARPA subterranean challenge," *arXiv preprint arXiv:2103.11470v1*, 2021.
- [16] L. Chang and K. Li, "Unified form for the robust gaussian information filtering based on M-estimate," *IEEE Signal Processing Letters*, vol. 24, no. 4, pp. 412–416, 2017.
- [17] R. G. Brown, P. Y. Hwang, et al., *Introduction to random signals and applied Kalman filtering*. Wiley New York, 1992, vol. 3.
- [18] A. Santamaria-Navarro, G. Loianno, J. Solà, V. Kumar, and J. Andrade-Cetto, "Autonomous navigation of micro aerial vehicles using high-rate and low-cost sensors," *Autonomous robots*, vol. 42, no. 6, pp. 1263–1280, 2018.
- [19] V. Madyastha, V. Ravindra, S. Mallikarjunan, and A. Goyal, "Extended Kalman filter vs. error state Kalman filter for aircraft attitude estimation," in *AIAA Guidance, Navigation, and Control Conference*, 2011, p. 6615.
- [20] M. Palieri, B. Morrell, A. Thakur, K. Ebadi, J. Nash, A. Chatterjee, C. Kanellakis, L. Carlone, C. Guaragnella, and A. Agha-mohammadi, "Locus: A multi-sensor lidar-centric solution for high-precision odometry and 3d mapping in real-time," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 421–428, 2020.