Circuits, Systems, and Signal Processing
https://doi.org/10.1007/s00034-020-01422-w

An Improved Tobit Kalman Filter with Adaptive Censoring Limits

Kostas Loumponias[1] · Nicholas Vretos[2] · George Tsaklidis[1] · Petros Daras[2]

Received: 7 February 2019 / Revised: 7 April 2020 / Accepted: 9 April 2020

Abstract

This paper deals with the Tobit Kalman filtering (TKF) process when the measurements are correlated and censored. The case of interval censoring, i.e., the case of measurements which belong to some interval with given censoring limits, is considered. Two improvements of the standard TKF process are proposed, in order to estimate the hidden state vectors. Firstly, the exact covariance matrix of the censored measurements is calculated by taking into account the censoring limits. Secondly, the probability of a latent (normally distributed) measurement to belong in or out of the uncensored region is calculated by taking into account the Kalman filter residual. The designed algorithm is tested using both synthetic and real data sets. The real data set includes human skeleton joints' coordinates captured by the Microsoft Kinect II sensor. In order to cope with certain real-life situations that cause problems in human skeleton tracking, such as (self)-occlusions, closely interacting persons, etc., adaptive censoring limits are used in the proposed TKF process. Experiments show that the proposed method outperforms other filtering processes in minimizing the overall root-mean-square error for synthetic and real data sets.

Keywords Censored data · Adaptive Tobit Kalman filter · Human skeleton tracking

✉ George Tsaklidis
tsaklidi@math.auth.gr

Kostas Loumponias
kostikasl@math.auth.gr

Nicholas Vretos
vretos@iti.gr

Petros Daras
daras@iti.gr

[1] Department of Mathematics, Aristotle University of Thessaloniki, 54124 Thessaloníki, Greece

[2] Information Technologies Institute, Centre for Research and Technology-Hellas, 6th km Charilaou, Thermi 57001, Thessaloníki, Greece

Published online: 08 May 2020

1 Introduction

Human skeleton motion tracking has been studied for several decades and remains a highly active research field due to its importance in several diverse domains like surveillance applications, medical applications, serious games, educational applications, high performance sports monitoring and others [27,41,44,45]. With the advent of commercial RGB-D sensors [7,36], human skeleton motion tracking has attracted a lot of attention due to the capacity of the sensors to reliably track skeletal joints. However, regardless of the significant progress that has been achieved in both sensors' development and human skeleton motion tracking research, many applications require more accurate tracking of the human skeleton position and motion. On the sensors' side, high performing sensors (such as the Vicon System), which are able to accurately track at high rates, are very expensive and cumbersome to deploy. On the other hand, affordable, commercial RGB-D solutions (i.e., the Microsoft Kinect, the Xtion Pro and others) often produce low-quality human skeleton motion tracking due to their inherent problems (low sampling frequency, moderate depth resolution, UV light interferences, etc.), and also due to their simplistic setup (usually only one such sensor is deployed, resulting in occluding areas and human self-occlusion).

To overcome these issues and provide an affordable and, at the same time, reliable solution to the human skeleton motion tracking task, research has been steered towards two general categories of methods: methods that exploit multiple RGB-D sensors [5,36] and methods that use various filters able to improve and smooth the sensors' measurements [6,9,10]. For the first category, two major flaws arise: 1) the increase in the cost for monitoring, capturing and processing, and 2) the interferences between devices, which add more noise and restrictions to the problem at hand, thus, making it harder to solve. For the latter, the main drawback is the lack of a framework able to provide reliable estimations of the human skeleton joints.

In this paper, a new method is introduced, which belongs to the second category of methods. The human skeleton motion tracking is improved by filtering the Kinect skeleton joints' measurements through a novel Kalman type filtering method adapted to restrictive conditions concerning human skeleton movements. The measurements that are corrected and filtered are the 25 Kinect's V2 skeletal joints, which are time series of 3D spatial coordinates in a 3D space centered in the physical center of the Kinect's infrared sensor.

In the literature, in order to filter spatial coordinates (or a signal), various filters, e.g., Kalman Filter (KF) [20,28], Savitzky–Golay filter (SGF) [38], Particle Filtering [4] and others have been proposed. One of the most common filters for signal filtering is KF, which provides optimal minimum mean square error (MMSE) estimates under the assumption that the state-space model is linear and the signal's measurements are normally distributed. However, KF may perform a poor filtering when the noisy signal contains some extreme measurements (outliers).

In the case where certain bounds of the denoised signal's values are considered, the extreme measurements can be treated by providing this information in the

KF process. In order to deal with that, the censored normal distribution in the KF estimation procedure [15,37] is introduced. The use of censored probabilities theory in data filtering was firstly introduced in [1], where the Tobit Kalman Filter (TKF) was proposed aiming to estimate an unknown state vector, $\mathbf{x}$, when censored measurements, $\mathbf{y}$, are present. In [29,30], TKF was utilized in order to filter spatial coordinates of human skeleton; however, no proofs for the TKF process were provided.

In this paper, a new filter is proposed, the so-called Adaptive Tobit Kalman Filter (ATKF), which considers an occluded or self-occluded Kinect's skeletal joint as a censored measurement. This filter takes advantage of the approaches presented in [1,30] and proposes a new proof. The proposed approach results in a more accurate estimation of the probability of a measurement to fall into the censoring region and as a consequence, it leads to a more accurate estimation of the state. The proposed ATKF also adapts its censoring region at each time step by considering previous states. The main contributions of this paper are:

1. A proof for accurately calculating the covariance matrix of the censored measurements in Tobit Kalman filtering, by incorporating the censoring limits into the equation of censored covariance.
2. A proof for accurately calculating the probabilities of a latent measurement, y^*, to belong in or out of the uncensored region, by taking into consideration the KF residual.
3. A new Adaptive Tobit Kalman Filter able to adapt the censoring limits at each time step.
4. As an application of contributions 1,2 and 3, a new method, which improves the human skeleton tracking in real-time applications is provided.
5. A new evaluation metric for human skeleton motion filtering to measure the performance of a filtering technique, when no ground truth data are available.

The rest of the paper is organized as follows. In Sect. 2, related works are described, while in Sect. 3, the proposed Adaptive Tobit Kalman Filter is presented in detail. In Sect. 4, experimental results are drawn, using artificial data as well as real human skeleton motion tracking data. Finally, Sect. 5, concludes the paper.

2 Related Work

Many approaches exist for filtering and motion tracking of the human skeleton either from images, videos or depth information. In this paper, only methods that are most relevant to the present work (based on data filtering) are mentioned. For a more detailed discussion, we refer to the books [12] and [35] for data filtering and human skeleton motion, respectively.

Similar to the proposed method, Microsoft [22] proposed various filters for filtering human skeleton motion data from Kinect devices. Two of them are the simple and the exponential moving average [38,42], but there is not any reference on how the time windows and the weights should be chosen, since these are application dependent. Edwards et al. [10] denoised human skeleton motion data (obtained by a Kinect V2

sensor) using four different filters: (1) the moving average, (2) KF, (3) the Holt double exponential filter [19] and (4) their proposed filter, consisting of a Kalman filter with a Wiener Process Acceleration (WPA) [43]. Both the averaging filter and KF had a good filtering performance, but they introduced relatively large amounts of latency, while the other two had good performance and low latency. Finally, the WPA Kalman filter exhibited the best overall performance.

Regarding the filtering process *per se*, the most known and well established filtering method is KF. In order to overcome several drawbacks of KF (mainly due to its linear nature), the extended Kalman Filter (EKF) was proposed in [23]. Although EKF is not characterized as an optimal estimator, it is an extension of the linear-system technique to a wider class of problems, which are nonlinear, as with most of the real-life problems. However, EKF tends to be unstable in many applications due to its local nature, leading to incorrect filtering of a signal that exhibits a high degree of nonlinearities. To overcome these problems, the unscented Kalman Filter (UKF) was proposed in [13,24]. UKF uses a deterministic sampling technique known as unscented transform [18] to gather a minimal set of points around a local mean. By doing so, it provides better results than EKF when the predict and the update functions are highly nonlinear and EKF has typically poor performance. Finally, a very successful method is the particle filtering [8], which is a Monte Carlo-based filtering method. Though particle filtering is generally very adaptable, it requires a high computational burden, making it practically unsuitable for many real-time applications.

In the area of censored statistics, all the above-mentioned methods have their drawbacks. In Allik [1], it is stated that the formulation of a standard KF, as an estimator for censored data, results in a biased estimation of the unknown state. EKF suffers from an undefined Jacobian at the censored region, resulting in an ill-posed Jacobian and thus exhibiting poor performance. On the other hand, UKF is a less computationally expensive approach than particle filtering; however, it is proven to be non-robust when the measurements are close to the censored region [1]. Furthermore, while particle filtering is suitable for estimating the state values when the measurements are censored in certain cases, it has a substantial computational cost. Finally, TKF provides unbiased, recursive estimates of the latent state variables in/near the uncensored regions. TKF is completely recursive and computationally inexpensive, making it a perfect candidate for real-time applications such as the human skeleton motion tracking. Nevertheless, TKF neither takes into account the censored area in calculating the censored measurements variance nor it adapts the limits of the censored area [3].

Fei Han et al. [16] concerned TKF for a class of linear discrete-time system with random parameters. The elements of the state space matrices are allowed to be random variables in order to reflect the reality. Furthermore, they established a novel weighting covariance formula to address the quadratic terms associated with the random matrices. Although their proposed method with only one censoring limit is coped.

In the area of human skeleton motion tracking, several methods have been proposed involving multiple RGB-D sensors, increasing the complexity and the cost of the solution as mentioned before. In [36], Sungphil et al. proposed a new method for human skeleton motion tracking using multiple Kinect V1 sensors. They determined the reliability of each 3D joint's position, by combining multiple observations based on Kinect measurements confidence (a value gathered from the sensor). They used the

variances of measurements noise in order to identify the contribution of an observation (i.e., a weight) to create a series of fused measurements. Furthermore, they explained how to estimate the variance of measurements noise for each joint through KF. Finally, they presented the average 3D position error of ten activities produced by: (1) their method, (2) a single Kinect and (3) a simple-average. In all activities but one (running), their method appeared to give better results than other methods compared with other methods provided in the paper.

3 Proposed Method

In this section, the censoring data theory and the well-known TKF [1] are briefly described in order to better highlight the proposed contributions. Then, an alternative approach to the classical TKF is demonstrated, where the update function is generated by taking into account the censoring limits in the measurements covariance matrix calculation, and thus, resulting in a more accurate evaluation of the censored measurements. Finally, ATKF for human skeleton motion tracking is introduced, where the censored region limits (boundaries) are not constant, as is the case in the standard TKF.

3.1 Censored and Truncated Data

Censoring is a condition in which the value of a measurement or observation is only partially known [40]. Censoring occurs when a value falls outside the range of a measuring instrument. For example, a bathroom scale might only measure up to 140 kg. If an 150 kg individual is weighed using that scale, the observer would only know that the individual's weight is at least 140 kg (partially known). Censoring should not be confused with the related idea of truncation; while by censoring, observations result either in knowing the exact value that applies or in knowing that the value lies into an interval, in the truncation case, only observations in a given range are considered by ignoring all the others. Different types of censoring exist [33], such as: left, right, interval, Type I and Type II censoring, respectively. In real-life problems, censored data are very frequent and to the best of our knowledge the concept of censoring in human skeleton motion tracking has never been used before.

3.2 Tobit Kalman Filters

As has been already stated, KF does not provide optimal or unbiased estimates for the states when the measurements are censored. This is due to the fact that the assumptions of KF [17] are not met when the noise measurements are censored. TKF [1,3,39] provides a classification scheme for all aforementioned types of censoring. The state-space model of TKF is defined as,

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{w}_k \tag{1}$$

$$\mathbf{y}_k^* = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \tag{2}$$

$$y_{k,i} = \begin{cases} y^*_{k,i}, & a_i < y^*_{k,i} < b_i \\ a_i & y^*_{k,i} \leq a_i \\ b_i, & y^*_{k,i} \geq b_i, \end{cases} \qquad i = 1, 2, \ldots, m \in \mathbb{N}, \tag{3}$$

where $\mathbf{x}_k$ is the state vector at discrete time step k, and $\mathbf{w}_k$ and $\mathbf{v}_k$ are random vectors following $N(\mathbf{0}, \mathbf{Q}_k)$ and $N(\mathbf{0}, \mathbf{R}_k)$, respectively, where $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbf{Q}_k$ and $\mathbf{R}_k$ are referred as the covariance matrices of the process and observation noise, respectively. $\mathbf{A}_k$ and $\mathbf{H}_k$ are the transition and the observation matrices, respectively, $\mathbf{y}_k = \{y_{k,i}\}_{i=1}^m, \mathbf{y}^*_k = \{y^*_{k,i}\}_{i=1}^m$ are the saturated observations (that are Left and Right censoring at the same time), and the latent observations, respectively, while $\mathbf{a} = \{a_i\}_{i=1}^m$ and $\mathbf{b} = \{b_i\}_{i=1}^m$ are the lower and upper limits of the uncensored region, respectively. Finally, m designates the dimensionality of the process (which is three in the case of 3D human skeleton motion data).

Next, the process of TKF is presented, where K in Algorithm 1 denotes the total number of the measurements. The vectors $\hat{\mathbf{x}}_k^-$ and $\hat{\mathbf{x}}_k$ denote the a priori and a posteriori state estimates at time step k, respectively, while $\mathbf{P}_k^-$ and $\mathbf{P}_k$ are the covariance matrices of the errors of the a priori and a posteriori state estimates, respectively. The covariance matrix, $Cov(\mathbf{y}_k|\mathbf{y}_{k-1})$, the mean vector, $\mathbb{E}(\mathbf{y}_k|\mathbf{y}_{k-1})$ of the censored measurement, $\mathbf{y}_k$, and the cross-covariance matrix $Cov(\mathbf{x}_k, \mathbf{y}_k|\mathbf{y}_{k-1})$ are provided in [1–3]. As can been stated by Algorithm 1, the process of TKF is recursive and evolves in two stages, the predict and the update stage, respectively. The predict function is the same as in case of standard KF [20], since the censored measurements are not used at this stage.

Algorithm 1 Standard Tobit Kalman Filter

$\mathbf{x}_0 \leftarrow \mathbf{0}_n$
$\mathbf{P}_0 \leftarrow \mathbf{0}_{n \times n}$
for `k=1:K` **do**

 # Predict Stage
$\hat{\mathbf{x}}_k^- \leftarrow \mathbf{A}_k \hat{\mathbf{x}}_{k-1}$
$\mathbf{P}_k^- \leftarrow \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{Q}_k$

 # Update Stage
$\mathbf{K}_k \leftarrow Cov(\mathbf{x}_k, \mathbf{y}_k|\mathbf{y}_{k-1}) \cdot Cov(\mathbf{y}_k|\mathbf{y}_{k-1})^{-1}$
$\hat{\mathbf{x}}_k \leftarrow \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbb{E}(\mathbf{y}_k|\mathbf{y}_{k-1}))$
$\mathbf{P}_k \leftarrow \mathbf{P}_k^- - \mathbf{K}_k \cdot Cov(\mathbf{x}_k, \mathbf{y}_k|\mathbf{y}_{k-1})^T$

end for

3.3 Censored Moments

In this section, the first moment, the second moment and the covariance of a censored measurement $\mathbf{y} = \{y_i\}_{i=1}^m$ (not truncated) with censoring limits $\mathbf{a} = \{a_i\}_{i=1}^m$ and

$\mathbf{b} = \{b_i\}_{i=1}^m$ are calculated. For that purpose, the following Proposition is needed [31]:

Proposition 1 *If the random variable* $\mathbf{y}^* = \{y_i^*\}_{i=1}^m$ *follows a* $m-$*dimensional normal distribution with density function* $f(\mathbf{y}^*)$*, mean value* $\boldsymbol{\mu} = \{\mu_i\}_{i=1}^m$ *and non-singular covariance matrix* $\boldsymbol{\Sigma} = \{\sigma_{i,j}\}_{i,j=1}^m$*, then the expected values of* y_i^* *and* $y_i^* \cdot y_j^*$ *given that* $a_k < y_k^* < b_k$, $k = 1, \ldots, m$*, are:*

$$\mathbb{E}(y_i^* | a_k < y_k^* < b_k, k = 1, \ldots, m) = \mu_i + \sum_{k=1}^{m} \sigma_{i,k} \big(F_k(a_k) - F_k(b_k)\big), \tag{4}$$

$$\begin{aligned} &\mathbb{E}(y_i^* y_j^* | a_k < y_k^* < b_k, k = 1, \ldots, m) \\ &\quad = \mu_i \cdot \mu_j + \sigma_{i,j} + \sum_{k=1}^{m} \sigma_{i,k} \frac{\sigma_{j,k}(a_k F_k(a_k) - b_k F_k(b_k))}{\sigma_{k,k}} \\ &\quad + \sum_{k=1}^{m} \sigma_{i,k} \sum_{q \neq i} \Big(\sigma_{j,q} - \frac{\sigma_{k,q}\sigma_{j,k}}{\sigma_{k,k}}\Big) \big[\big(F_{k,q}(a_k, a_q) - F_{k,q}(a_k, b_q)\big) \\ &\quad - \big(F_{k,q}(b_k, a_q) - F_{k,q}(b_k, b_q)\big)\big]. \end{aligned} \tag{5}$$

The functions $F_i(x)$ and $F_{i,j}(x, y)$ are given by:

$$F_i(x) = \frac{1}{P(a_j < y_j^* < b_j, j = 1, \ldots, m)} \cdot \int_{a_1}^{b_1} \cdots \int_{a_{i-1}}^{b_{i-1}} \int_{a_{i+1}}^{b_{i+1}} \cdots \int_{a_m}^{b_m} f(x, \mathbf{y}_{-i}^*) \mathbf{dy}_{-i}^*, \tag{6}$$

$$\begin{aligned} F_{i,j}(x, y) = &\frac{1}{P(a_j < y_j^* < b_j, j = 1, \ldots, m)} \cdot \int_{a_1}^{b_1} .. \int_{a_{i-1}}^{b_{i-1}} \int_{a_{i+1}}^{b_{i+1}} .. \int_{a_{j-1}}^{b_{j-1}} \int_{a_{j+1}}^{b_{j+1}} .. \\ &\int_{a_m}^{b_m} f(x, y, \mathbf{y}_{-i-j}^*) \mathbf{dy}_{-i-j}^*, \end{aligned} \tag{7}$$

where $\mathbf{y}_{-i}^* = (y_1^*, .., y_{i-1}^*, y_{i+1}^*, .., y_m^*)$ and $\mathbf{y}_{-i-j}^* = (y_1^*, .., y_{i-1}^*, y_{i+1}^*, .., y_{j-1}^*, y_{j+1}^*, .., y_m^*)$. Next, the following Lemma is provided in order to calculate the censored moments.

Lemma 1 *Let* x *be a continuous random variable on a probability space* Ω*,* z *a discrete random variable with outcomes* $\{z_i\}_{i=1}^n$ *and* $f(x, z)$ *the joint probability function of* (x, z)*. Then, the expected value of* (x, z) *can be given by*

$$\mathbb{E}(x,z) = \sum_{i=1}^{n} z_i \, E(x|z=z_i) P(z=z_i). \tag{8}$$

Proof

$$\begin{aligned}
\mathbb{E}(x,z) &= \int_{\Omega} \sum_{i=1}^{n} z_i x f(x,z) \mathrm{d}x \\
&= \int_{\Omega} \sum_{i=1}^{n} z_i x f(x|z_i) P(z=z_i) \mathrm{d}x \\
&= \sum_{i=1}^{n} z_i P(z=z_i) \int_{\Omega} x f(x|z_i) \mathrm{d}x \\
&= \sum_{i=1}^{n} z_i P(z=z_i) \, \mathbb{E}(x|z=z_i).
\end{aligned}$$

□

Now, the following Proposition can be proved (see "Appendix A") using Lemma 1 and Proposition 1:

Proposition 2 *The mean value of the censored variable y_i with censoring limits a_i and b_i can be written as:*

$$\begin{aligned}
\mathbb{E}(y_i) = \mu_i P(a_i < y_i^* < b_i) + \sigma_{i,i}(f_i(a_i) - f_i(b_i)) \\
+ a_i P(y_i^* \le a_i) + b_i P(y_i^* \ge b_i).
\end{aligned} \tag{9}$$

It is noted that the mean value of the censored variable, y_i, depends on the censoring limits a_i, b_i, and not on the censoring limits of the others components y_j for $j \neq i$. Furthermore, it can be proved (see "Appendix B") that:

Proposition 3 *The variance and the joint mean value of the censored variable y_i and y_i, y_j, respectively, with censoring limits $\{a_i, b_i\}$ and $\{a_i, b_i, a_j, b_j\}$, respectively, are given by:*

$$\begin{aligned}
Var(y_i) = {} & \mu_i^2 (1 - P_{un}^i) P_{un}^i + \sigma_{i,i} P_{un}^i + a_i^2 (1 - P_a^i) P_a^i \\
& + b_i^2 (1 - P_b^i) P_b^i - 2 a_i b_i P_a^i P_b^i - \sigma_{i,i}^2 (f(a_i) - f(b_i)) \\
& + 2 \mu_i \sigma_{i,i} (f_i(a_i) - f(b_i))(1 - P_{un}^i) \\
& + \sigma_{i,i} \big((a_i - \mu_i) f_i(a_i) - (b_i - \mu_i) f_i(b_i) \big) \\
& - 2 \Big(\mu_i P_{un}^i + \sigma_{i,i} \big(f_i(a_i) - f(b_i) \big) \Big) \Big(a_i P_a^i + b_i P_b^i \Big)
\end{aligned} \tag{10}$$

and

$$\begin{aligned}\mathbb{E}(y_i y_j) = {} & a_i b_j P(1) + b_i b_j P(3) + a_i a_j P(7) + b_i a_j P(9) \\ & + b_j\, \mathbb{E}(y_i^* | a_i < y_i^* < b_i, y_j^* \geq b_j) P(2) \\ & + a_i\, \mathbb{E}(y_j^* | a_j < y_j^* < b_j, y_i^* \leq a_i) P(4) \\ & + \mathbb{E}(y_i^* y_j^* | a_i < y_i^* < b_i, a_j < y_j^* < b_j) P(5) \\ & + b_i\, \mathbb{E}(y_j^* | a_j < y_j^* < b_j, y_i^* \geq b_i) P(6) \\ & + a_j\, \mathbb{E}(y_i^* | a_i < y_i^* < b_i, y_j^* \leq a_j) P(8).\end{aligned} \tag{11}$$

The probabilities P_{un}^i, P_a^i, P_b^i and $P(j)_{j=1,\ldots,9}$ are defined as follows:

$$\begin{aligned}& P_{un}^i = P(a_i < y_i^* < b_i),\ P_a^i = P(y_i^* \leq a_i), \\ & P_b^i = P(y_i^* \geq b_i),\ P(1) = P(y_i^* \leq a_i, y_j^* \geq b_j), \\ & P(2) = P(a_i < y_i^* < b_i, y_j^* \geq b_j), \\ & P(3) = P(y_i^* \geq b_i, y_j^* \geq b_j), \\ & P(4) = P(y_i^* \leq a_i, a_j < y_j^* < b_j), \\ & P(5) = P(a_i < y_i^* < b_i, a_j < y_j^* < b_j), \\ & P(6) = P(y_i^* \geq b_i, a_j < y_j^* < b_j), \\ & P(7) = P(y_i^* \leq a_i, y_j^* \leq a_j), \\ & P(8) = P(a_i < y_i^* < b_i, y_j^* \leq a_j), \\ & P(9) = P(y_i^* \geq b_i, y_j^* \leq a_j).\end{aligned}$$

The truncated expected values $\mathbb{E}(y_i^* | \cdot)$, $\mathbb{E}(y_i^* y_j^* | \cdot)$ in (11) are calculated in "Appendix B". Hence, the covariance matrix of the censored variable $\mathbf{y}$ can be calculated by (9)-(11). It is noted that $V(y_i)$ and $\mathbb{E}(y_i y_j)$ depend on the censoring limits a_i, b_i and a_i, b_i, a_j, b_j, respectively, and not on the censoring limits of the others components. This property does not hold in the case of truncated moments (5).

3.4 Corrected Tobit Kalman Filter

In this paper, as in [1–3,16], the a posteriori estimation, $\hat{\mathbf{x}}_k$, is calculated as a linear combination of the a priori estimation, $\hat{\mathbf{x}}_k^-$, and the censored measurement $\mathbf{y}_k$ (see Algorithm 1). Although these estimations are not optimal, it is proved that they minimize the trace of state error covariance [32]. Next, the changes made by the proposed TKF are provided .

The cross-covariance matrix $\mathbf{R}_{k,1} = Cov(\mathbf{x}_k, \mathbf{y}_k | \mathbf{y}_{k-1})$ has been calculated in [2,3] and takes the form

$$\mathbf{R}_{k,1} = \mathbf{P}_k^- \mathbf{H}_k^T \mathbf{P}_{un,k}, \tag{12}$$

where $\mathbf{P}_{un,k}$ is a $m \times m$ diagonal matrix, and its entries are the probabilities of a measurement to be uncensored, at time step k. More specifically, the ith diagonal element of $\mathbf{P}_{un,k}$, is the probability that a latent measurement $y^*_{k,i}$ belongs to the uncensored region. Furthermore, the entries of the diagonal matrices $\mathbf{P}_{\mathbf{a},k}$, $\mathbf{P}_{\mathbf{b},k}$ denote the probabilities of a measurement to be censored from below or above, respectively, at time step k. It is proved (see "Appendix C") that:

$$\mathbf{P}_{un,k} = diag\begin{bmatrix} \Phi(b_{k,1}) - \Phi(a_{k,1}) \\ \dots \\ \Phi(b_{k,m}) - \Phi(a_{k,m}) \end{bmatrix}, \tag{13}$$

$$\mathbf{P}_{\mathbf{a},k} = diag\begin{bmatrix} \Phi(a_{k,1}) \\ \dots \\ \Phi(a_{k,m}) \end{bmatrix}, \tag{14}$$

$$\mathbf{P}_{\mathbf{b},k} = diag\begin{bmatrix} 1 - \Phi(b_{k,1}) \\ \dots \\ 1 - \Phi(b_{k,m}) \end{bmatrix}, \tag{15}$$

where Φ stands for the cumulative function of $N(0, 1)$. The amounts $b_{k,i}$ and $a_{k,i}$ are calculated as ("Appendix C")

$$b_{k,i} = \frac{b_i - m_{k,i}}{\sqrt{s_{(i,i),k}}} \tag{16}$$

$$a_{k,i} = \frac{a_i - m_{k,i}}{\sqrt{s_{(i,i),k}}}, \tag{17}$$

where $\mathbf{S}_k = \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k + \mathbf{R}_k = \{s_{(i,j),k}\}$ and $\mathbf{m}_k = \mathbf{H}\hat{\mathbf{x}}_k^- = \{m_{k,i}\}$. In standard TKF [1–3], these amounts are calculated as (which are denoted with $*$ to not confuse them with the proposed)

$$b^*_{k,i} = \frac{b_i - m_{k,i}}{\sqrt{r_{(i,i),k}}} \tag{18}$$

$$a^*_{k,i} = \frac{a_i - m_{k,i}}{\sqrt{r_{(i,i),k}}}, \tag{19}$$

where $\mathbf{R}_k = \{r_{(i,j),k}\}$ is the covariance matrix of observation (measurement) noise.

It is worth to mention that in (18) and (19) the information from the KF residual, $(\mathbf{y}^*_k - \mathbf{m}_k)$, is omitted. More specifically, in (16) and (17), as opposed to (18) and (19), the term $(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k)_{i,i}$ is incorporated in the denominator, which consequently, adds information into (16) and (17), concerning the KF residual. By doing so, the probability of a measurement to belong to the uncensored region is estimated more accurately.

The mean vector of the censored measurement $\mathbf{y}_k$ given the previous censored measurement $\mathbf{y}_{k-1}$ can be written (in matrix notation) using (9) as:

$$\mathbb{E}(\mathbf{y}_k|\mathbf{y}_{k-1}) = \mathbf{m}_k \cdot \mathbf{P}_{un,k} + \mathbf{S}_k \cdot diag(f_i(a_i) - f_i(b_i))_{i=1}^m + \mathbf{a} \cdot \mathbf{P}_{\mathbf{a},k} + \mathbf{b} \cdot \mathbf{P}_{\mathbf{b},k}. \tag{20}$$

The entries of the censored covariance matrix, $\mathbf{R}_{k,2} = Cov(\mathbf{y}_k|\mathbf{y}_{k-1})$, given the last censored measurement, $\mathbf{y}_{k-1}$, are equal with

$$(R_{k,2})_{i,j} = \begin{cases} Var(y_{k,i}|\mathbf{y}_{k-1}),\ i = j \\ \mathbb{E}(y_{k,i}\, y_{k,j}|\mathbf{y}_k) - \mathbb{E}(y_{k,i}|\mathbf{y}_k)\, \mathbb{E}(y_{k,j}|\mathbf{y}_k),\ i \neq j \end{cases}. \tag{21}$$

In particular, the diagonal elements of $\mathbf{R}_{k,2}$ are calculated as $Var(y_i)$ (10), where the mean vector, $\boldsymbol{\mu}$, and covariance matrix, $\boldsymbol{\Sigma}$, in the proposed model are equal with $\mathbf{m}_k = \mathbf{H}\hat{\mathbf{x}}_k^-$ and $\mathbf{S}_k = \mathbf{H}\mathbf{P}_k^-\mathbf{H} + \mathbf{R}_k$, respectively, and the probabilities $P^i_{un,k}$, $P^i_{a,k}$, $P^i_{b,k}$ for $i = 1, \ldots, m$ are given in (13)–(15). In the same way, the off-diagonal elements of $\mathbf{R}_{k,2}$ are calculated as $\mathbb{E}(y_i\, y_j)$ (11). In what follows, TKF^c denotes the proposed method, where the censored moments are calculated through (12), (20) and (21).

In [2], the covariance matrix, $\mathbf{R}^*_{k,2}$, of the censored measurement $\mathbf{y}_k$, is given by

$$\mathbf{R}^*_{k,2} = \mathbf{P}^*_{un,k}\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T\mathbf{P}^*_{un,k} + \mathbf{R}^*_k, \tag{22}$$

where $\mathbf{R}^*_k$ and $\mathbf{P}^*_{un,k}$ are diagonal matrices, where their entries are the truncated variances of $y^*_{k,i}$ (3) and the probabilities $\left(\Phi(b^*_{k,i}) - \Phi(a^*_{k,i})\right)$ for $i = 1, \ldots, m$, respectively.

The main difference between (21) and (22) is that in (22) the limits a_i and b_i appear only in $\mathbf{P}^*_{un,k}$. In the case where $a_i = 0$ and b_i is big enough (that is, only non-negative measurements are considered), then (22) provides a satisfactory approximation of the covariance matrix of the censored measurements. In order to clarify the notation and illustrate the difference between $\mathbf{R}_{k,2}$ and $\mathbf{R}^*_{k,2}$, an illustrative example is provided as follows: the censored covariance matrix for the random multidimensional $\mathbf{Y}^* \sim N(\mathbf{m}_k, \mathbf{S}_k)$ is examined, with censoring limits $\mathbf{a} = (-1, -3, 1)^t$ and $\mathbf{b} = (1, 7, 4)^t$. The mean vector, $\mathbf{m}_k$, and the covariance matrix $\mathbf{S}_k$ are defined to be equal with,

$$\mathbf{m}_k = \mathbf{H}_k \cdot (2, 2, 3)^T,$$

and

$$\mathbf{S}_k = \mathbf{H}_k \begin{bmatrix} 4 & 3 & 4 \\ 3 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix} \mathbf{H}_k^T + \mathbf{R}_k,$$

while, without loss of generality, $\mathbf{H}_k$ and $\mathbf{R}_k$ are defined to be equal with the 3×3 identity matrix. Then, the process is as follows: 1) 10^5 random measurements are produced by $N(\mathbf{m}_k, \mathbf{S}_k)$ 100 times. 2) Each time, the sampling covariance matrix derived from the censored measurements is calculated. 3) The arithmetic mean of $\mathbf{R}_s$, of the 100 sampling covariance matrices is calculated. 4) The covariance matrices $\mathbf{R}_{k,2}$

and $\mathbf{R}^*_{k,2}$ are calculated by (21) and (22), respectively. As it can be seen by (23)-(25), the proposed covariance matrix, $\mathbf{R}_{k,2}$, is almost identical with the sampling covariance matrix, $\mathbf{R}_s$.

$$\mathbf{R}_s = \begin{bmatrix} 0.4648 & 0.6962 & 0.5083 \\ 0.6962 & 4.7754 & 1.9195 \\ 0.5083 & 1.9195 & 1.4384 \end{bmatrix}, \tag{23}$$

$$\mathbf{R}_{k,2} = \begin{bmatrix} 0.4651 & 0.6962 & 0.5085 \\ 0.6962 & 4.7747 & 1.9189 \\ 0.5085 & 1.9189 & 1.4379 \end{bmatrix}, \tag{24}$$

$$\mathbf{R}^*_{k,2} = \begin{bmatrix} 0.2724 & 0.4719 & 0.5151 \\ 0.4719 & 5.0000 & 3.2744 \\ 0.5151 & 3.2744 & 3.2002 \end{bmatrix}. \tag{25}$$

The marginal probability function, $f(y_{k,i}|\mathbf{y}_{k-1})$, of the ith component of the censored measurement $\mathbf{y}_k$ given the last measurement, $\mathbf{y}_{k-1}$, is,

$$\begin{aligned} f(y_{k,i}|\mathbf{y}_{k-1}) = {} & \frac{1}{\sqrt{s_{(i,i),k}}}\phi\left(\frac{y_{k,i}-m_{k,i}}{\sqrt{s_{(i,i),k}}}\right)u(y_{k,i}-a_i)u(b_i-y_{k,i}) \\ & + \Phi(a_{k,i})\delta(a_i-y_{k,i}) + (1-\Phi(b_{k,i}))\delta(b_i-y_{k,i}), \end{aligned} \tag{26}$$

where ϕ and Φ are the probability and the cumulative distribution functions of the standard normal distribution, respectively, δ is the Kronecker delta function and u stands for the Heavyside function, where $u(x) = 1$, when $x > 0$ and $u(x) = 0$, otherwise.

The next step in our procedure is to calculate the likelihood function by taking into consideration the censored data distribution. The likelihood function for the censored measurements $\{y_{k,i}\}_{k=1}^K$ by (26), (14) and (15) can be calculated as:

$$\begin{aligned} L_i(y_{1,i},\ldots,y_{K,i}) = {} & \prod_{y_{k,i}=a_i}\Phi(a_{k,i}) \times \prod_{y_{k,i}=b_i}(1-\Phi(b_{k,i})) \\ & \times \prod_{a_i<\,y_{k,i}<\,b_i}\frac{1}{\sqrt{s_{(i,i),k}}}\phi\left(\frac{y_{k,i}-m_{k,i}}{\sqrt{s_{(i,i),k}}}\right). \end{aligned} \tag{27}$$

In the case that the components of $\mathbf{y}_k$ are mutually independent, the likelihood function of the censored measurements $\{\mathbf{y}_k\}_{k=1}^K$ takes the form:

$$L(\mathbf{y}_1,\ldots,\mathbf{y}_K) = \prod_{i=1}^m L_i(y_{1,i},\ldots,y_{K,i}). \tag{28}$$

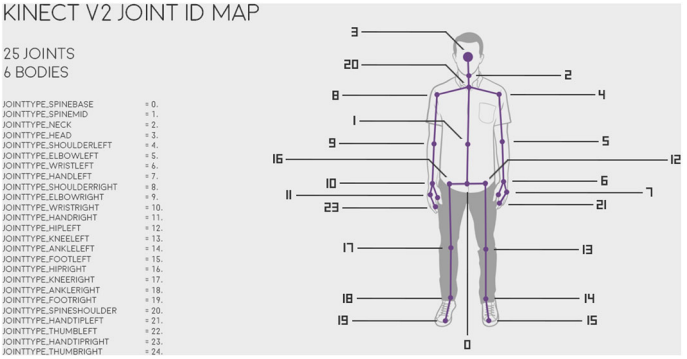


Fig. 1 Human skeleton's joints map of the Kinect V2 sensor

In the case of [1], the likelihood function becomes

$$L_i^*(y_{1,i},\ldots,y_{K,i}) = \prod_{y_{k,i}=a_i} \Phi(a_{k,i}^*) \times \prod_{y_{k,i}=b_i} \left(1-\Phi(b_{k,i}^*)\right) \times \prod_{a_i < y_{k,i} < b_i} \frac{1}{\sqrt{r_{(i,i),k}}} \phi\left(\frac{y_{k,i}-m_{k,i}}{\sqrt{r_{(i,i),k}}}\right). \tag{29}$$

Note that the denominator does not take into account the specific distribution of the measurements.

3.5 Adaptive Tobit Kalman Filter used to Human Skeleton Tracking

In what follows, the Microsoft Kinect V2 sensor is utilized to record 3D point sequences (human skeletons) of a human in motion [21]. In human skeleton tracking, the body is represented by a number of joints (25 in total), corresponding to different body parts such as head, neck, shoulders, etc (see Fig. 1). Each joint is represented by the vector of its Euclidean 3D space coordinates (z_1, z_2, z_3) and the aim is to denoise the measurements for every joint in order to improve the representation of human movements. Thus, each one of the joints' coordinates is denoised separately; the input is the vector of the joints' coordinates, $\mathbf{y}_k^* = (y_{k,1}^*, y_{k,2}^*, y_{k,3}^*)$ (latent measurement), and the output is the vector of the denoised states coordinates, $\mathbf{x}_k = (x_{k,1}, x_{k,2}, x_{k,3})$.

To start tracking, the initial observation, $\mathbf{H}_k$, and the transition, $\mathbf{A}_k$, matrices are defined to be equal to the identity matrix. Therefore, the covariance matrix of the noise measurement, $\mathbf{R}_k$ is defined to be

$$\mathbf{R}_k = 0.01 \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{30}$$

$\mathbf{R}_k$ is chosen to initialize in that way, under the assumption that Kinect exhibits significant errors in human skeleton tracking. To support that claim, small scale experiments

are conducted, proving that even if a person is at rest and in front of the Kinect, the error in the displacement estimation between measurement and ground truth data is almost 0.02 m [11,34], thus a variance of 0.01 m^2 seems to be a valid choice.

In KF [6,10], no restrictions in joints' movements have been taken into account, as opposed to the proposed method. To that end, beyond the Kinect V2 sensor, the state-of-the-art Vicon tracking system has been used as a ground truth reference. In Vicon data, for various recordings, it is observed that the velocity of the spatial coordinates z_1 and z_3 did not exceed 34 cm per frame, for every joint, and the z_2 coordinate did not exceed 18 cm per frame. In what follows, these restrictions are used in order to correct the data produced by the Kinect sensor. By applying these restrictions, ATKF is constructed with limits $\mathbf{l}_k^1$ and $\mathbf{l}_k^2$ for the vector of the spatial coordinates, $(y_{k,1}^*, y_{k,2}^*, y_{k,3}^*)$, as follows:

$$\mathbf{l}_k^2 = \mathbf{H}_k \hat{\mathbf{x}}_{k-1} + \mathbf{c}, \tag{31}$$

and

$$\mathbf{l}_k^1 = \mathbf{H}_k \hat{\mathbf{x}}_{k-1} - \mathbf{c}, \tag{32}$$

where the observation matrix, $\mathbf{H}_k$, is the identity matrix, $\mathbf{l}_k^1$ and $\mathbf{l}_k^2$ are the limits of ATKF at time k, which depend on the previous estimation of spatial coordinates, $\hat{\mathbf{x}}_{k-1}$, and the vector $\mathbf{c}$, which for human skeleton tracking is experimentally found to be

$$\mathbf{c} = (0.34, 0.18, 0.34). \tag{33}$$

Thus, for the latent measurement $\mathbf{y}_k^* = (y_{k,1}^*, y_{k,2}^*, y_{k,3}^*)$ at time k, it arises

$$y_{k,i} = \begin{cases} y_{k,i}^*, & l_{k,i}^1 < y_{k,i}^* < l_{k,i}^2 \\ l_{k,i}^1, & y_{k,i}^* \le l_{k,i}^1 \\ l_{k,i}^2, & y_{k,i}^* \ge l_{k,i}^2. \end{cases} \qquad i = 1, 2, 3. \tag{34}$$

In Algorithm 2, the ATKF procedure for human skeleton tracking is summarized.

This model corrects Kinect measurements, when they have high abnormal velocity. It should be noted that, if $l_{k,i}^1 \to -\infty$ and $l_{k,i}^2 \to \infty$ (i.e., the range of ATKF limits becomes very large), ATKF tends to the standard KF, because in this case the Kinect measurements belong to the uncensored region and consequently they are known. Due to this fact, in some recordings, which do not include big or fast joints' movements, (thus, the Kinect measurements always belong to the uncensored region) are expected to get almost the same results concerning RMSE for KF as well as for ATKF.

In order to create a general model for filtering Kinect V2 measurements without having to estimate the matrix $\mathbf{Q}_k$ for every time-window (because this is time consuming), it is assumed that this matrix is constant. Substituting for $\mathbf{R}_k$ in the likelihood function (27), the covariance matrix of the noise process, $\mathbf{Q}_k$, can be estimated. By

Algorithm 2 ATKF for Human Skeleton Tracking

$\mathbf{x}_0 \leftarrow \mathbf{0}_n$
$\mathbf{P}_0 \leftarrow \mathbf{0}_{n\times n}$
for k=1:K **do**

  if k==1 **then**
    Use the standard KF
  else if then
    #Calculate the censoring limits
    $\mathbf{l}_k^2 \leftarrow \mathbf{H}_k\hat{\mathbf{x}}_{k-1} + \mathbf{c}$
    $\mathbf{l}_k^1 \leftarrow \mathbf{H}_k\hat{\mathbf{x}}_{k-1} - \mathbf{c}$

    #Censored measurements
    $\mathbf{y}_k \leftarrow \mathbf{l}_k^1 \cdot (\mathbf{y}_k^* \le \mathbf{l}_k^1) + \mathbf{l}_k^2 \cdot (\mathbf{y}_k^* \ge \mathbf{l}_k^2) + \mathbf{y}_k^* \cdot (\mathbf{l}_k^1 < \mathbf{y}_k^* < \mathbf{l}_k^2)$

    #Predict Stage
    $\hat{\mathbf{x}}_k^- \leftarrow \mathbf{A}_k\hat{\mathbf{x}}_{k-1}$
    $\mathbf{P}_k^- \leftarrow \mathbf{A}_k\mathbf{P}_{k-1}\mathbf{A}_k^T + \mathbf{Q}_k$

    #Update Stage
    $\mathbf{R}_{k,1} \leftarrow$ (12)
    $\mathbb{E}(\mathbf{y}_k|\mathbf{y}_{k-1}) \leftarrow$ (20)
    $\mathbf{R}_{k,2} \leftarrow$ (21)
    $\mathbf{K}_k \leftarrow \mathbf{R}_{k,1} \cdot \mathbf{R}_{k,2}^{-1}$
    $\hat{\mathbf{x}}_k \leftarrow \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbb{E}(\mathbf{y}_k|\mathbf{y}_{k-1}))$
    $\mathbf{P}_k \leftarrow \mathbf{P}_k^- - \mathbf{K}_k \cdot \mathbf{R}_{k,1}^T$

  end if
end for

experimenting on various joints' movements, it is derived that the values of $\mathbf{Q}_k$ are smaller than those of matrix $\mathbf{R}_k$ and generally they depend on the speed of the human skeleton's joints. Regarding slow joints' movements, the entries of $\mathbf{Q}_k$ are smaller than 10^{-4} and for faster joints' movements they lie between 10^{-3} and 10^{-2}. In some cases, where the entries of $\mathbf{Q}_k$ appeared to be quite large (in the order of 10^{-2}), the human skeleton moved too quickly in an abnormal manner due to occlusions and/or self-occlusions. Thereafter, the covariance matrix of the noise process is assumed to be equal with:

$$\mathbf{Q}_k = 0.0025 \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{35}$$

otherwise, if smaller or larger values are defined, ATKF will be either over-smoothed or will not denoise the Kinect measurements. Therefore, the matrix $\mathbf{Q}_k$ given in (35), seems to be an appropriate choice for filtering the Kinect V2 sensor measurements of human skeleton tracking.

4 Experiments

In this section, three sets of experiments are conducted to evaluate TKF^c and ATKF compared to other methods. 1) TKF and 2) TKF^c are utilized in the first experimental set (oscillator), which is employed in [2]. Next, 1) SGF, 2) KF, 3) TKF, 4) TKF^c and 5) ATKF are used in order to denoise data for two different experimental sets: a) Real-life data captured by a Kinect sensor, b) Real-life data captured by both a Kinect sensor and a Vicon system.

4.1 Oscillator

In the first experimental set, a motivating example of tracking a sinusoidal model by a TKF and TKF^c is presented, when the measurements are saturated. Let the state space equations have the form of (3), with state space matrices

$$\mathbf{A} = c \cdot \begin{bmatrix} \cos(w) & -\sin(w) \\ \sin(w) & cos(w) \end{bmatrix}, \tag{36}$$

and

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \tag{37}$$

where $c = 0.999$ and $w = 0.005 \cdot 2\pi$. The disturbance $\mathbf{w}_k$ is assumed to be normally distributed, i.e., $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q})$, where

$$\mathbf{Q} = \begin{bmatrix} 0.05^2 & 0 \\ 0 & 0.05^2 \end{bmatrix}, \tag{38}$$

while, the measurement noise, v_k, is normally distributed, $v_k \sim N(0, 0.5)$. The initial state vector is equal to $\mathbf{x}_0 = [5 \quad 0]^T$ with covariance matrix $\mathbf{P}_0 = \mathbf{I}_2$, the censored limits are $a = -0.5$ and $b = 0.5$. Therefore, by the above example censored (saturated) measurements, y_k, are produced where $k = 1, \ldots, K = 1000$.

Next, the above process is repeated $M_s = 100$ times and the filters' RMSEs are calculated for each iteration. The means of filterer's RMSEs for 100 iterations are presented in Table 1, where separate RMSEs for the two estimated coordinates of the state vector, $\mathbf{x}_k$, are provided. It can be observed that the corrected TKF^c outperforms TKF in state estimation (Fig. 2). This is due to the fact that in TKF some important terms are ignored when calculating $\mathbf{R}^*_{k,2}$ (22), while these terms are included in TKF^c process (21).

In addition to the RMSE assessment, the proposed method is evaluated by the noncredibility index (NCI) [25]. In [25,26], it has been shown that NCI is free of serious drawbacks of other metrics such as the average normalized estimation error squared [26] (ANESS). Furthermore, it is concluded that an estimator is not credible if NCI is significantly greater than 1. Finally, NCI for M_s Monte Carlo simulations at time index k is defined as

Table 1 The mean of RMSEs for the filters TKF and TKF^c, respectively

Filter	Mean RMSE of $\hat{x}_1$	Mean RMSE of $\hat{x}_2$
TKF	0.4434	0.5464
TKF^c	**0.4066**	**0.5192**

$$NCI(k) = \frac{10}{M_s}\sum_{j=1}^{M_s}\left| log_{10}\frac{\left(\mathbf{x}_k(j)-\hat{\mathbf{x}}_k(j)\right)^T\mathbf{P}_k^{-1}(j)\left(\mathbf{x}_k(j)-\hat{\mathbf{x}}_k(j)\right)}{\left(\mathbf{x}_k(j)-\hat{\mathbf{x}}_k(j)\right)^T\mathbf{P}_k^{*-1}\left(\left(\mathbf{x}_k(j)-\hat{\mathbf{x}}_k(j)\right)\right.}\right| \tag{39}$$

where $\hat{\mathbf{x}}_k(j)$ and $\mathbf{P}_k(j)$ are the a posteriori estimate of $\mathbf{x}_k(j)$ and the covariance matrix of the error of the a posteriori estimate at j-th Monte Carlo simulation, respectively, while the matrix $\mathbf{P}_k^*$ is given by

$$\mathbf{P}_k^* = \frac{1}{M_s}\sum_{j=1}^{M_s}\left(\mathbf{x}_k(j)-\hat{\mathbf{x}}_k(j)\right)\left(\mathbf{x}_k(j)-\hat{\mathbf{x}}_k(j)\right)^T. \tag{40}$$

The arithmetic average of filters' NCI for all estimates is presented in Table 2. It can been observed that the proposed method outperforms TKF in state estimation. As it can be seen in Fig. 3, the $NCI(k)$ indices are in most cases smaller for TKF^c than for TKF.

4.2 Recordings by the Kinect Sensor

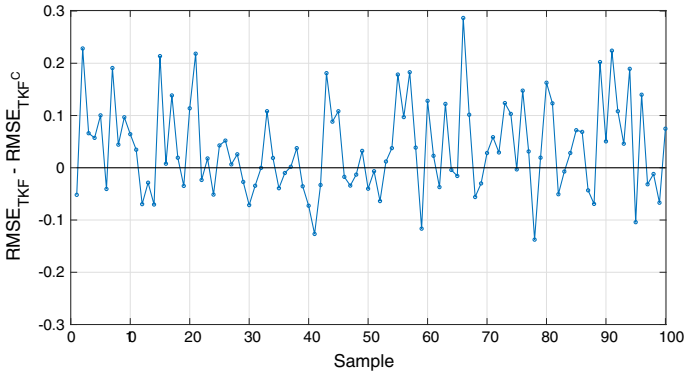
In the second experiment set, various human movements are recorded by a single Kinect V2 sensor. In some of the recordings, the human skeleton motion exhibits an important error on the z_2 axis (practically, the human skeleton seems to "fall down") for one or two frames. The above-mentioned filters are applied to correct this specific error.

In order to evaluate the performance of the different filters, a novel metric, m_i, is proposed to better examine the result of filtering the joints' movements. Let us denote by $g_{k,i}$ the filtering of the ith component of the measurement $y_{k,i}$ at time k. Then,

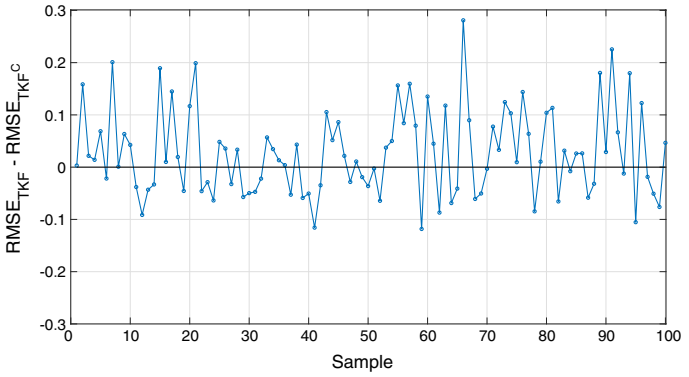
$$m_i = \text{average}\left[(dg_{k,i})^2\right]_{k=1}^{n-1}, \tag{41}$$

where $dg_{k,i} = g_{k+1,i} - g_{k,i}$ and n is the number of measurements.

In the case of TKF^c and TKF, the device limits are used. For instance, the ranges of Kinect spatial coordinates z_1 (width), z_2 (height) and z_3 (depth) are approximately $[-3\,\text{m},\ 3\text{m}]$, $[-1.5\,\text{m},\ 3\,\text{m}]$ (if the Kinect V2 sensor is located 1.5 m over the ground) and $[0.5\,\text{m},\ 5\,\text{m}]$, respectively. Thus, these limits for the Kinect measurements are used in order to test TKF and TKF^c. The covariance matrices of TKF^c and TKF for the noise measurement, $\mathbf{R}$, are defined as in ATKF (30), while the covariance matrices for the noise process, $\mathbf{Q}$, can be estimated using the likelihood functions (29) and (28), respectively. By experimenting on various joints' movements, it results that the entries



(a) $\hat{x}_1$



(b) $\hat{x}_2$

Fig. 2 The difference between TKF's and TKF^c's RMSE for each iteration

Table 2 The mean of RMSEs for the filters TKF and TKF^c, respectively

Filter	Mean of NCI
TKF	1.1760
TKF^c	**0.9898**

of **Q** are the same as in the case of ATKF; therefore, the same matrix **Q** is used given by (35). In the case of KF, the covariance matrix, **R**, is defined as in ATKF (30) and the covariance matrix for the noise process, **Q**, is estimated by the log-likelihood function given in [14]. The results showed (in the same experiments as it is mentioned earlier), that the entries of **Q** are almost the same as in the case of ATKF; thus, the matrix **Q** is defined as in (35).

In the experiments, the overall average M of the metrics m_i for various recordings is calculated. The results showed that ATKF achieves better performance in noise reduction than the other filters (see Table 3), especially in the cases where the skeleton seems to collapse, while KF, TKF^c and TKF have almost the same overall average M and SGF has a poor performance. As can be seen in Fig. 4 for two different experiments,

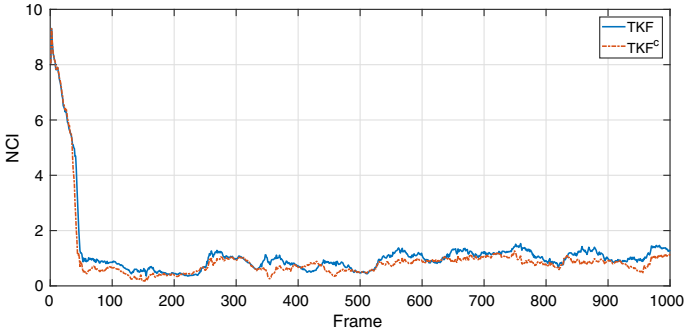


Fig. 3 NCI for the filters TKF and TKF^c, respectively

Table 3 The overall average M of the recordings for the Kinect V2 sensor and the filters

Filter	Overall Average M
SGF	0.790×10^{-3}
KF	0.436×10^{-3}
TKF	0.433×10^{-3}
TKF ^{c}	0.433×10^{-3}
ATKF	$\mathbf{0.362 \times 10^{-3}}$
Kinect V2	1.70×10^{-3}

the head's spatial coordinates z_2 of the human skeleton resulted from ATKF do not (correctly) follow the error produced by the Kinect sensor. It can be seen (Fig. 4) that although KF, TKF^c and TKF improve the human skeleton motion, they provide inferior results than the ones produced by ATKF, while SGF has the worst performance among all. In the first experiment illustrated in Fig. 4a, the ATKF skeleton followed the sharp "fall" for almost 5 cm, while KF, TKF^c and TKF skeletons for 12 cm, and the SGF skeleton for 20 cm. The joint-based average m_i as opposed to the overall experiments average M of ATKF in this experiment is 0.350×10^{-3}, while in KF, TKF^c and TKF are $0.409 * 10^{-3}$ and in SGF is 0.797×10^{-3}.

To better illustrate the superiority of ATFK, the motion of the human skeleton (obtained by Kinect) under heavy occlusion is illustrated in the first row of subfigures in Fig. 5 for four consecutive frames. The first subfigure shows the human skeleton one frame before "collapsing," the next two show the human skeleton under heavy occlusion and the last one shows a better performance of human skeleton. In the next five rows of Fig. 5, the motion of human skeleton is illustrated as it is resulted by SGF, KF, TKF, TKF^c and ATKF, respectively. All filters had a delay of 1–2 frames due to the occluded area, but ATKF clearly outperforms all other methods (see the last row in Fig. 5)

4.3 Recording by Kinect Sensor and Vicon System

In this subsection, the proposed method with respect to ground truth data is evaluated. To that end, an athlete throwing a ball with his right hand is monitored, and this motion

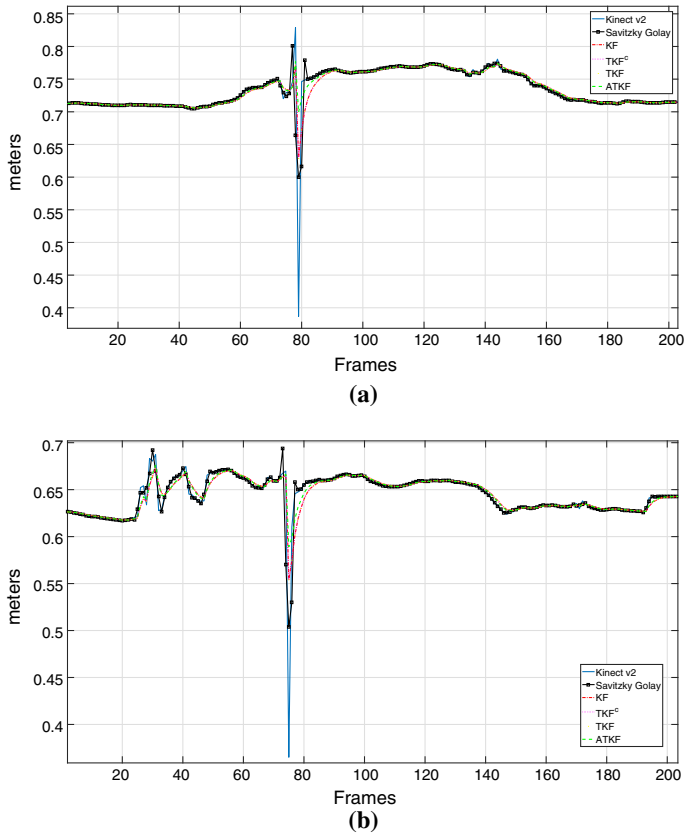


Fig. 4 The head's spatial coordinates y_k of Kinect V2 sensor, Savitzky–Golay, KF, TKF, TKF^c and ATKF

is recorded by a Kinect V2 sensor and the Vicon system at the same time. Vicon is used as the ground truth in order to compare results using the proposed method on Kinect measurements. The number of Kinect's and Vicon's frames are 266 (almost 8.8667 s) and 139 (4.4480 s), respectively. The Kinect time-stamp is almost 0.033 s per frame, while Vicon time-stamp is constantly 0.032 s. Vicon data are interpolated in order to deal with the time-stamp problem; after interpolation, the new Vicon data include 133 frames. Therefore, the two sensors are temporally synchronized to start together. To do so, initially the angles of knees and elbows obtained by Kinect and Vicon data are calculated and then, the RMSEs between these angles for different delays are calculated. The results show that the minimum values of RMSE for every angle appeared for delays of 92–95 frames. The different delays between the angles in some cases are somewhat expected because Kinect records fast movements with delay (i.e., after some frames).

It is noticed that KF filters the spatial coordinates without affecting the movement (see Fig. 6). TKF^c and TKF perform exactly the same filtering in all joints as KF, while SGF does not perform a satisfactory filtering in some points where the measurements have a significant error. Table 4 gives the RMSEs for the angles as they arise for

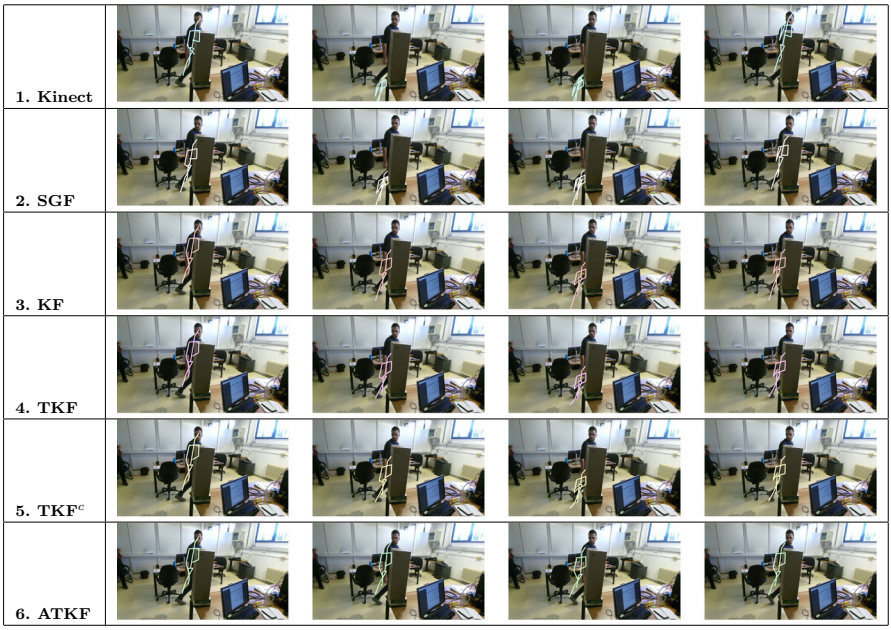
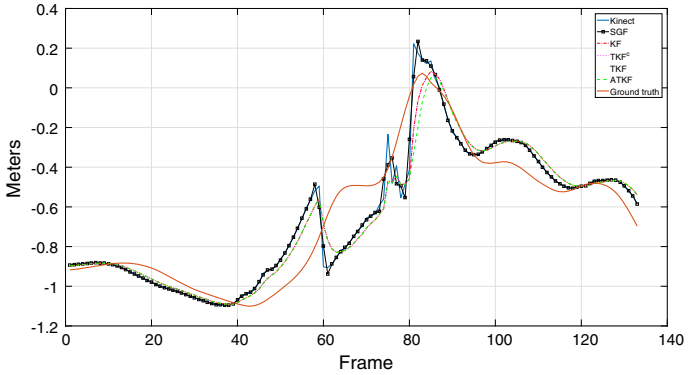


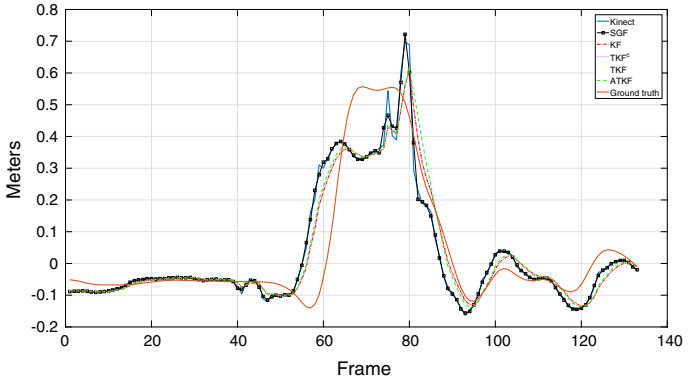
Fig. 5 Each row represents the human skeleton motion for four consecutive frames as it is obtained by (1) Kinect V2 sensor, (2) SGF, (3) KF, (4) TKF, (5) TKF^c and (6) ATKF, respectively

Table 4 RMSEs for the angles by Kinect V2, SGF, KF, TKF, TKF^c and ATKF for time delay 92, 93, 94 and 95

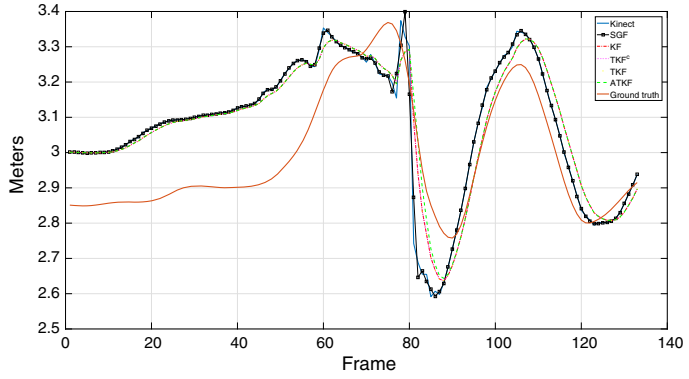
Angles	Kin. v2	SGF	KF	TKF	TKF^c	**ATKF**
Right elbow	39.31	37.44	36.60	36.60	36.60	**36.32**
Left elbow	31.58	30.65	27.98	27.98	27.98	**26.50**
Right knee	16.70	16.79	15.79	15.79	15.79	**14.90**
Left knee	26.25	25.81	25.14	25.14	25.14	**25.11**
Right elbow	38.76	36.86	35.90	35.90	35.90	**35.57**
Left elbow	32.18	31.27	28.43	28.43	28.43	**27.02**
Right knee	17.03	17.12	15.75	15.75	15.75	**14.93**
Left knee	26.38	26.01	24.85	24.85	24.85	**24.82**
Right elbow	38.43	36.63	35.40	35.40	35.40	**35.06**
Left elbow	32.99	32.09	29.08	29.08	29.08	**27.75**
Right knee	17.77	17.79	16.04	16.04	16.04	**15.26**
Left knee	26.67	26.46	24.90	24.90	24.90	**24.89**
Right elbow	38.39	36.64	35.25	35.25	35.25	**34.93**
Left elbow	33.96	33.06	29.92	29.92	29.92	**28.70**
Right knee	18.78	18.78	16.58	16.58	16.58	**15.77**
Left knee	27.14	27.02	25.24	25.24	25.24	**25.23**



(a) The x_k coordinates of the right hand.



(b) The y_k coordinates of the right hand.



(c) The z_k coordinates of the right hand

Fig. 6 The right hand's coordinates by Kinect V2 sensor, SGF, KF, TKF^c, TKF, ATKF and Ground truth

delays $t = 92, 93, 94, 95$ frames, respectively. In all cases, the RMSEs are big enough because of the occlusion of some joints during the recording.

In Fig. 6, the right hand's coordinates resulted by KF, TKF, TKF^c and ATKF are almost the same, because all measurements belong to the uncensored region, while SGF coordinates are almost the same with Kinect's coordinates. However, as can be seen in Table 4 , in all cases concerning RMSEs, better results are achieved via ATKF compared to those of standard KF, TKF^c and TKF. The RMSEs of SGF are almost the same as the Kinect RMSEs.

5 Conclusion and Discussion

The aim of this paper is to improve (1) the well-known TKF process [1] and (2) the human skeleton motion tracking using a single Kinect V2 sensor, which often generates noisy measurements due to occlusion, lighting conditions, etc. To that end, a novel filtering method, called ATKF, was proposed which relies on the censored data statistics theory for human skeleton motion tracking in real-time. In order to estimate the hidden state vector by the censored measurement, firstly, the probabilities of a latent measurement to belong in or out of the uncensored region were evaluated ("Appendix C") and secondly, the accurate covariance matrix of the censored normal distribution ("Appendix B") was evaluated. In this approach, the limits of the uncensored region had to be defined for the Kinect's measurements, in a reasonable manner for every time step k. To do so, many data with various joints movements were tested, which were obtained by ground truth sensor, such as the Vicon tracking system.

The proposed method was evaluated against (1) standard KF, (2) TKF, (3) TKF^c with constant limits and 4) SGF in three different setups: (1) Artificial data (2) Kinect and (3) Kinect plus Vicon human skeleton motion data. A new metric was introduced in order to evaluate results when no ground truth is available. Finally, the covariance matrix, $\mathbf{Q}$, of the noise process was calculated under a specific experimental methodology as opposed to previous methods where random or simple experimental covariance matrices were used. Among the five approaches, ATKF gave better results in all the different setups for human skeleton tracking.

In a future work, it would be interesting to use the proposed filtering method for action recognition tasks in the wild, where uncontrolled environments and situations where RGB-D sensors may have poor performance often occur. Moreover, as a step beyond, it would be interesting to consider the state vector $\mathbf{x}$ as a censored state, aiming at achieving a more accurate filtering of the human skeleton motion data.

Acknowledgements This work was supported by the European Project (Horizon2020) ICT4Life under GA 690090.

Appendix A: The Censored Mean Value

In what follows, the proof of Proposition 2 is provided:

Proof of Proposition 2, Sect. 3.3 For a discrete random variable $z_i \sim B(p_i)$ (Bernoulli distribution) in Lemma 1, it is derived that

$$\mathbb{E}(x, z_i) = \mathbb{E}(x|z_i = 1) \cdot p_i. \tag{42}$$

The censored measurement, y_i, can be written in terms of Bernoulli distributions; therefore, the censored mean value is written by (42) as,

$$\mathbb{E}(y_i) = \sum_{j=1}^{3^{n-1}} \mathbb{E}(y_i^*|a_i < y_i^* < b_i, R_j) P((a_i, b_i), R_j) + a_i P(y_i^* \le a_i) + b_i P(y_i^* \ge b_i), \tag{43}$$

where the first term is the sum of all possible mean values of $\mathbb{E}(y_i|a_i < y_i < b_i)$ given that the rest variables lie in a region $R_j = [(L_1, U_1), \ldots, (L_{i-1}, U_{i-1}), (L_{i+1}, U_{i+1}), \ldots, (L_n, U_n)]$, where

$$(L_k, U_k) = \begin{cases} (-\infty, a_k) & or \\ (a_k, b_k) & or \\ (b_k, \infty) \end{cases}$$

where j=1,…,3^{n-1}. $P((a_i, b_i), R_j)$ denotes the probability of variable $\mathbf{y}^*$ to lie in a region $[(a_i, b_i), R_j]$. It is derived by (5) that

$$\begin{aligned} \mathbb{E}(y_i) &= \sum_{j=1}^{3^{n-1}} \Big(\mu_i + \sum_{k=1}^{n} \sigma_{i,k} \big(F_k(L_k) - F_k(U_k)\big)_{R_j} \Big) P(R_j) + a_i P(y_i^* \le a_i) \\ &\quad + b_i P(y_i^* \ge b_i) \\ &= \sum_{k=1}^{n} \sum_{j=1}^{3^{n-1}} \sigma_{i,k} \big(F_k(L_k) - F_k(U_k)\big)_{R_j} P(R_j) + \mu_i P(a_i < y_i^* < b_i) \\ &\quad + a_i P(y_i^* \le a_i) + b_i P(y_i^* \ge b_i), \end{aligned} \tag{44}$$

where $\big(F_k(L_k) - F_k(U_k)\big)_{R_j}$ is the difference of functions (6) given that the variable $\mathbf{y}^*$ lies in the region $R_j \cup (a_i, b_i)$. In the case where $k \neq i$, it is derived that:

$$\begin{aligned} \sum_{j=1}^{3^{n-1}} \sigma_{i,k} \big(F_k(L_k) - F_k(U_k)\big)_{R_j} P(R_j) &= \sum_{j=1}^{3^{n-2}} \sigma_{i,k} \big(F_k(-\infty) - F_k(a_k)\big)_{V_j} P(V_j) \\ &\quad + \sum_{j=1}^{3^{n-2}} \sigma_{i,k} \big(F_k(a_k) - F_k(b_k)\big)_{V_j} P(V_j) \\ &\quad + \sum_{j=1}^{3^{n-2}} \sigma_{i,k} \big(F_k(b_k) - F_k(\infty)\big)_{V_j} P(V_j) = 0, \end{aligned} \tag{45}$$

where V_j is the region

$$[(L_1, U_1), \ldots, (L_{k-1}, U_{k-1}), (L_{k+1}, U_{k+1}), \ldots, (a_i, b_i), \ldots, (L_n, U_n)].$$

In the case where $k = i$, it is derived that,

$$\begin{aligned}
&\sum_{j=1}^{N} \sigma_{i,k}\big(F_k(L_k) - F_k(U_k)\big)_{R_j} P(R_j) \\
&\quad = \sum_{j=1}^{N} \sigma_{i,i}\big(F_i(a_i) - F_i(b_i)\big)_{R_j} P(R_j) \\
&\quad = \sigma_{i,i}\big(f_i(a_i) - f_i(b_i)\big),
\end{aligned} \tag{46}$$

where $f_i(y_i^*)$ is the normal distribution of $y_i^* \sim N(\mu_i, \sigma_{i,i})$. Thus, by (44)-(46) arises

$$\mathbb{E}(y_i) = \mu_i P(a_i < y_i^* < b_i) + \sigma_{i,i}(f_i(a_i) - f_i(b_i)) + a_i P(y_i^* \le a_i) + b_i P(y_i^* \ge b_i). \tag{47}$$

□

Appendix B: The Censored Covariance Matrix

In what follows, the proof of Proposition 3 is provided:

Proof of Proposition 3, Sect. 3.3 In the same way as for the censored mean ("Appendix A"), it is proved that the second moment of y_i depends on the censoring limits $\{a_i, b_i\}$. Therefore, it is derived by Lemma 1 that

$$\mathbb{E}\left(y_i^2\right) = \mathbb{E}(y_i^{*2} | a_i < y_i^* < b_i) P(a_i < y_i^* < b_i) + a_i^2 P(y_i^* \le a_i) + b_i^2 P(y_i^* \ge b_i),$$

where the first term [31] is equal with

$$\begin{aligned}
\mathbb{E}(y_i^{*2} | a_i < y_i^* < b_i) = \sigma_{i,i} + \mu_i^2 + 2\mu_i \sigma_{i,i} \frac{f_i(a_i) - f(b_i)}{P(a_i < y_i^* < b_i)} \\
+ \sigma_{i,i} \frac{(a_i - \mu_i) f_i(a_i) - (b_i - \mu_i) f_i(b_i)}{P(a_i < y_i^* < b_i)}.
\end{aligned} \tag{48}$$

Therefore, it is derived by (48) that,

$$\begin{aligned}
\mathbb{E}\left(y_i^2\right) &= (\sigma_{i,i} + \mu_i^2) P(a_i < y_i^* < b_i) \\
&\quad + \sigma_{i,i}\big((a_i - \mu_i) f_i(a_i) - (b_i - \mu_i) f_i(b_i)\big) \\
&\quad + 2\mu_i \sigma_{i,i}(f_i(a_i) - f(b_i)) + a_i^2 P(y_i^* \le a_i) + b_i^2 P(y_i^* \ge b_i).
\end{aligned} \tag{49}$$

Finally, the censored variance is given by

$$\begin{aligned} Var(y_i) = {} & \mu_i^2(1-P_{un}^i)P_{un}^i + \sigma_{i,i}P_{un}^i + a_i^2(1-P_a^i)P_a^i \\ & + b_i^2(1-P_b^i)P_b^i - 2a_ib_iP_a^iP_b^i - \sigma_{i,i}^2(f(a_i)-f(b_i)) \\ & + 2\mu_i\sigma_{i,i}(f_i(a_i)-f(b_i))(1-P_{un}^i) \\ & + \sigma_{i,i}\big((a_i-\mu_i)f_i(a_i)-(b_i-\mu_i)f_i(b_i)\big) \\ & - 2\Big(\mu_iP_{un}^i + \sigma_{i,i}\big(f_i(a_i)-f(b_i)\big)\Big)\Big(a_iP_a^i + b_iP_b^i\Big), \end{aligned} \tag{50}$$

where $P_{un}^i = P(a_i < y_i^* < b_i)$, $P_a^i = P(y_i^* \le a_i)$ and $P_b^i = P(y_i^* \ge b_i)$. The expectation value of $y_i \cdot y_j$ is written by Lemma 1 as:

$$\begin{aligned} \mathbb{E}(y_iy_j) = {} & a_ib_jP(1) + b_ib_jP(3) + a_ia_jP(7) + b_ia_jP(9) \\ & + b_j\sum_{k=1}^{3^{n-2}}\mathbb{E}(y_i^*|a_i < y_i^* < b_i, y_j^* \ge b_j, G_k)P(G_k) \\ & + a_i\sum_{k=1}^{3^{n-2}}\mathbb{E}(y_j^*|a_j < y_j^* < b_j, y_i^* \le a_i, G_k)P(G_k) \\ & + \sum_{k=1}^{3^{n-2}}\mathbb{E}(y_iy_j^*|a_i < y_i^* < b_i, a_j < y_j^* < b_j, G_k)P(G_k) \\ & + b_i\sum_{k=1}^{3^{n-2}}\mathbb{E}(y_j^*|a_j < y_j^* < b_j, y_i^* \ge b_i, G_k)P(G_k) \\ & + a_j\sum_{k=1}^{3^{n-2}}\mathbb{E}(y_i^*|a_i < y_i^* < b_i, y_j^* \le a_j, G_k)P(G_k), \end{aligned} \tag{51}$$

where

$$\begin{aligned} P(1) = P(y_i^* \le a_i, y_j^* \ge b_j),\ P(3) = P(y_i^* \ge b_i, y_j^* \ge b_j), \\ P(7) = P(y_i^* \le a_i, y_j^* \le a_j),\ P(9) = P(y_i^* \ge b_i, y_j^* \le a_j), \end{aligned}$$

and G_k for $k = 1, \ldots, 3^{n-2}$ denote a region (as in the case of the censored mean) where the multi-variable, $\mathbf{y}_{-i-j}^*$, lies on.

Concerning the last five terms of (51), it is proved (as in case of second moment) that they depend only on the censoring limits $\{a_i, b_i, a_j, b_j\}$; thus, (51) can be written

as

$$\begin{aligned}\mathbb{E}(y_i y_j) = {} & a_i b_j P(1) + b_i b_j P(3) + a_i a_j P(7) + b_i a_j P(9) \\ & + b_j \, \mathbb{E}(y_i^* | a_i < y_i^* < b_i, y_j^* \geq b_j) P(2) \\ & + a_i \, \mathbb{E}(y_j^* | a_j < y_j^* < b_j, y_i^* \leq a_i) P(4) \\ & + \mathbb{E}(y_i^* y_j^* | a_i < y_i^* < b_i, a_j < y_j^* < b_j) P(5) \\ & + b_i \, \mathbb{E}(y_j^* | a_j < y_j^* < b_j, y_i^* \geq b_i) P(6) \\ & + a_j \, \mathbb{E}(y_i^* | a_i < y_i^* < b_i, y_j^* \leq a_j) P(8),\end{aligned} \tag{52}$$

where

$$\begin{aligned}P(2) &= P(a_i < y_i^* < b_i, y_j^* \geq b_j), \\ P(4) &= P(y_i^* \leq a_i, a_j < y_j^* < b_j), \\ P(5) &= P(a_i < y_i^* < b_i, a_j < y_j^* < b_j), \\ P(6) &= P(y_i^* \geq b_i, a_j < y_j^* < b_j), \\ P(8) &= P(a_i < y_i^* < b_i, y_j^* \leq a_j).\end{aligned}$$

At this point, it should be noted that the truncated moments $\mathbb{E}(y_i^* | \cdot)$ and $\mathbb{E}(y_i^* y_j^* | \cdot)$ in (52) are calculated by (5) and (5), respectively. Although the functions (6), (7) in our case (censoring measurements) are defined only for the variables y_i^* and y_j^*, i.e.,:

$$F_i(x) = \frac{\int_{a_j}^{b_j} f_{Y_i^*, Y_j^*}(x, y_j^*) dy_j^*}{P(a_j < y_j^* < b_j, a_i < y_i^* < b_i)},$$

and

$$F_{i,j}(x, y) = \frac{f_{Y_i^*, Y_j^*}(x, y)}{P(a_j < y_j^* < b_j, a_i < y_i^* < b_i)}.$$

Therefore, the covariance matrix can be defined by the terms (47), (50) and (52). □

Appendix C: Evaluation of the Probabilities of the Latent Measurement to Belong to the Censored or Uncensored Region

In what follows, the proofs for (13)–(15) are provided.

The mean of the latent measurement $\mathbf{y}_k^*$ given the saturated measurement $\mathbf{y}_{k-1}$ is

$$\mathbf{m}_k = \mathbb{E}(\mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k | \mathbf{y}_{k-1}) = \mathbf{H}_k \, \mathbb{E}(\mathbf{x}_k | \mathbf{y}_{k-1}) = \mathbf{H}_k \hat{\mathbf{x}}_k^- . \tag{53}$$

The covariance matrix of $\mathbf{y}_k^* - \mathbf{H}_k\hat{\mathbf{x}}_k^-$ is

$$\begin{aligned}\mathbf{S}_k = \mathrm{Cov}(\mathbf{y}_k^* - \mathbf{H}_k\hat{\mathbf{x}}_k^-) &= \mathrm{Cov}(\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-) \\ &= \mathrm{Cov}\big(\mathbf{H}_k(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)\big) + \mathrm{Cov}(\mathbf{v}_k)\end{aligned}$$

thus,

$$\mathbf{S}_k = \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k. \tag{54}$$

By (53) and (54), it is clear that $\mathbf{y}_k^*|\mathbf{y}_{k-1} \sim N(\mathbf{m}_k, \mathbf{S}_k)$. The probability $P^i_{\mathbf{a},k}$ of the ith component of the latent measurement $\mathbf{y}_k^*$ to be equal or less than a_i is

$$\begin{aligned}P^i_{\mathbf{a},k} = P(y^*_{k,i} \le a_i) &= P\left(\frac{y^*_{k,i} - m_{k,i}}{\sqrt{s_{(i,i),k}}} \le \frac{a_i - m_{k,i}}{\sqrt{s_{(i,i),k}}}\right) \\ &= \Phi\left(\frac{a_i - m_{k,i}}{\sqrt{s_{(i,i),k}}}\right).\end{aligned} \tag{55}$$

Following the same procedure, the probability $P^i_{\mathbf{b},k}$ of the ith component of the latent measurement $\mathbf{y}_k^*$ to be equal or bigger than b_i is

$$P^i_{\mathbf{b},k} = 1 - \Phi\left(\frac{b_i - m_{k,i}}{\sqrt{s_{(i,i),k}}}\right). \tag{56}$$

Finally, the probability of the ith component of the latent measurement $\mathbf{y}_k^*$ to lie in the uncensored region (a_i, b_i) is

$$P^i_{un,k} = 1 - P^i_{\mathbf{a},k} - P^i_{\mathbf{b},k}. \tag{57}$$

□

References

1. B. Allik, The tobit Kalman filter: an estimator for censored data. Ph.D. thesis, University of Delaware (2014)
2. B. Allik, C. Miller, M.J. Piovoso, R. Zurakowski, Estimation of saturated data using the tobit Kalman filter, in *2014 American Control Conference* (IEEE) (2014), pp. 4151–4156
3. B. Allik, C. Miller, M.J. Piovoso, R. Zurakowski, The tobit Kalman filter: An estimator for censored measurements. IEEE Trans. Control Syst. Technol. **24**(1), 365–371 (2016)
4. M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. IEEE Trans. Signal Process. **50**(2), 174–188 (2002)
5. S. Asteriadis, A. Chatzitofis, D. Zarpalas, D.S. Alexiadis, P. Daras, Estimating human motion from multiple kinect sensors, in *Proceedings of the 6th International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications* (ACM) (2013), p. 3
6. E.M. Berti, A.J.S. Salmerón, F. Benimeli, Kalman filter for tracking robotic arms using low cost 3d vision systems, in *The Fifth International Conference on Advances in Computer–Human Interactions* (2012), pp. 236–240

7. F. Destelle, A. Ahmadi, N.E. O'Connor, K. Moran, A. Chatzitofis, D. Zarpalas, P. Daras, Low-cost accurate skeleton tracking based on fusion of kinect and wearable inertial sensors, in *2014 22nd European Signal Processing Conference (EUSIPCO)* (IEEE) (2014), pp. 371–375
8. J. Deutscher, A. Blake, I. Reid, Articulated body motion capture by annealed particle filtering, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2000.* vol. 2 (IEEE) (2000), pp. 126–133
9. Y. Du, W. Wang, L. Wang, Hierarchical recurrent neural network for skeleton based action recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 1110–1118
10. M. Edwards, R. Green, Low-latency filtering of kinect skeleton data for video game control, in *Proceedings of the 29th International Conference on Image and Vision Computing New Zealand* (ACM) (2014), pp. 190–195
11. B. Galna, G. Barry, D. Jackson, D. Mhiripiri, P. Olivier, L. Rochester, Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease. Gait Posture **39**(4), 1062–1068 (2014)
12. M.S. Grewal, *Kalman Filtering* (Springer, Berlin, 2011)
13. F. Gustafsson, G. Hendeby, Some relations between extended and unscented Kalman filters. IEEE Trans. Signal Process. **60**(2), 545–555 (2012)
14. J.D. Hamilton, *Time Series Analysis*, vol. 2 (Princeton University Press, Princeton, 1994)
15. J. Hampshire, J.W. Strohbehn, Tobit maximum-likelihood estimation for stochastic time series affected by receiver saturation. IEEE Trans. Inf. Theory **38**(2), 457–469 (1992)
16. F. Han, H. Dong, Z. Wang, G. Li, F.E. Alsaadi, Improved tobit Kalman filtering for systems with random parameters via conditional expectation. Signal Process. **147**, 33–45 (2018)
17. A.C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, Cambridge, 1990)
18. S.J. Julier, The scaled unscented transformation, in *Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301)*, vol. 6, (IEEE) (2002), pp. 4555–4559
19. P.S. Kalekar, Time series forecasting using holt-winters exponential smoothing. Kanwal Rekhi Sch. Inf. Technol. **4329008**, 1–13 (2004)
20. R.E. Kalman, A new approach to linear filtering and prediction problems. J. Basic Eng. **82**(1), 35–45 (1960)
21. A. Kar, Skeletal tracking using microsoft kinect. Methodology **1**, 1–11 (2010)
22. M. Kinect, Skeletal joint smoothing white paper. https://msdn.microsoft.com/en-us/library/jj131429.aspx/. Accessed July 07 2016
23. B.F. La Scala, R.R. Bitmead, Design of an extended Kalman filter frequency tracker. IEEE Trans. Signal Process. **44**(3), 739–742 (1996)
24. A.B.L. Larsen, S. Hauberg, K.S. Pedersen. Unscented Kalman filtering for articulated human tracking, in *Scandinavian Conference on Image Analysis* (Springer) (2011), pp. 228–237
25. X.R. Li, Z. Zhao, Measuring estimator's credibility: noncredibility index, in *2006 9th International Conference on Information Fusion* (IEEE) (2006), pp. 1–8
26. X.R. Li, Z. Zhao, V.P. Jilkov, Estimator's credibility and its measures, in *Proceedings of the IFAC 15th World Congress* (2002)
27. Z. Li, Z. Wei, Y. Yue, H. Wang, W. Jia, L.E. Burke, T. Baranowski, M. Sun, An adaptive hidden markov model for activity recognition based on a wearable multi-sensor device. J. Med. Syst. **39**(5), 1–10 (2015)
28. T.J. Lim, Y. Ma, The Kalman filter as the optimal linear minimum mean-squared error multiuser cdma detector. IEEE Trans. Inf. Theory **46**(7), 2561–2566 (2000)
29. K. Loumponias, N. Vretos, P. Daras, G. Tsaklidis, Using tobit Kalman filtering in order to improve the motion recorded by microsoft kinect, in *Proceedings of the International workshop on applied probability (IWAP), Toronto, Canada* (2016)
30. K. Loumponias, N. Vretos, G. Tsaklidis, P. Daras, Using Kalman filter and tobit Kalman filter in order to improve the motion recorded by kinect sensor ii, in *Proceedings of the 29th Panhellenic Statistics Conference, Naousa, Greece* (2016), pp. 322–334
31. B. Manjunath, S. Wilhelm, Moments calculation for the double truncated multivariate normal density (2009)
32. C. Masreliez, R. Martin, Robust bayesian estimation for the linear model and robustifying the Kalman filter. IEEE Trans. Autom. Control **22**(3), 361–371 (1977)

33. R.G. Miller Jr., *Survival Analysis*, vol. 66 (Wiley, Hoboken, 2011)
34. A. Mobini, S. Behzadipour, M. Saadat Foumani, Accuracy of kinect's skeleton tracking for upper body rehabilitation applications. Disabil Rehabil Assist Technol **9**(4), 344–352 (2014)
35. T.B. Moeslund, A. Hilton, V. Krüger, L. Sigal, *Visual Analysis of Humans* (Springer, Berlin, 2011)
36. S. Moon, Y. Park, D.W. Ko, I.H. Suh, Multiple kinect sensor fusion for human skeleton tracking using Kalman filtering. Int. J. Adv. Robot. Syst. **13**, 65 (2016)
37. P. Moore, The estimation of the mean of a censored normal distribution by ordered variables. Biometrika **43**(3/4), 482–485 (1956)
38. A. Savitzky, M.J. Golay, Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. **36**(8), 1627–1639 (1964)
39. J. Tobin, Estimation of relationships for limited dependent variables. Econom. J. Econom. Soc. **26**(1), 24–36 (1958)
40. B.W. Turnbull, The empirical distribution function with arbitrarily grouped, censored and truncated data. J. R. Stat. Soc. Ser. B (Methodol.) **38**, 290–295 (1976)
41. J. Wang, Z. Liu, Y. Wu, J. Yuan, Learning actionlet ensemble for 3d human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. **36**(5), 914–927 (2014)
42. W.W.-S. Wei, *Time Series Analysis* (Addison-Wesley publ Reading, Boston, 1994)
43. G. Whitmore, F. Schenkelberg, Modelling accelerated degradation data using wiener diffusion with a time scale transformation. Lifetime Data Anal. **3**(1), 27–45 (1997)
44. X. Yang, Y. Tian, Super normal vector for human activity recognition with depth cameras. IEEE Trans. Pattern Anal. Mach. Intell. **PP**(99), 1–1 (2016)
45. C. Zhang, L. Zhang, Activity recognition in smart homes based on second-order hidden markov model. Int. J. Smart Home **7**(6), 237–244 (2013)