Greek Statistical Institute
Proceedings of the 29th Panhellenic Statistics Conference (2016), pp. 000−000

USING KALMAN FILTER AND TOBIT KALMAN FILTER IN ORDER TO IMPROVE THE MOTION RECORDED BY KINECT SENSOR II

K. Loumponias[1,2], N. Vretos[1], P. Daras[1], G. Tsaklidis[2]
[1]Centre for Research and Technology Hellas, Information Technologies Institute, Thessaloniki, Greece.
{loumponias, vretos, daras}@iti.gr
[2]Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki, Greece.
{kostikasl, tsaklidi}@math.auth.gr

ABSTRACT

In this paper, two methods are proposed to analyse skeleton data recorded by the Kinect v2 sensor using Kalman filter and Tobit Kalman filter in order to minimize the noise of the acquisition device due to occlusions, self occlusions e.t.c. The skeleton data are three-dimensional spatial coordinates that record movements of an individual's joints. The variance of the noise process is estimated using the likelihood function. In order to include into the model restrictive conditions based on the joints displacements per frame, we apply the Tobit Kalman Filter. Experiments on skeleton data show that the Tobit Kalman filter corrects better the noise than the Kalman filter.

Keywords: Kinect Sensor, Skeleton Data, Standard Kalman Filter, Tobit Kalman Filter.

1. INTRODUCTION

Skeleton tracking motion is a scientific area studied by means of depth sensors (i.e., sensors using the depth coordinate as basic coordinate), proved to be very useful in many applications, such as monitoring of activity recognition [Zhen et al (2015), Zhang et al (2013)] and health tracking [Galna et al (2014)]. In the present paper it is shown that the Kinect 2 sensor is able to achieve skeleton tracking performance in a low-cost manner for activity recognition; this sensor is able to track at most 25 joints.
However, Kinect sensor generates a lot of noise because of self-occlusion and lack of accuracy in fast movements. Especially when the skeleton's joints are occluded, they often appear to be shifted in a no reasonable manner.
The method presented in this paper is based on appropriately smoothing the joints' spatial coordinates. In order to smooth the coordinates, various common stochastic filters are used, e.g. filters based on moving average, Savitzky-Golay

filter e.t.c. [Yong et al (2015)], without any restrictions in joints' movements. Now we want to develop a model which will not allow the joints to move abnormally, and without affect the real movements. For this reason, we studied the joints' speeds by carrying out various experiments using groundtruth sensors. Then we applied the Tobit Kalman filter by taking into account the speeds restrictions.
In [Sungphil et al (2016)] a method based on the use of multiple Kinect sensors for skeleton tracking is proposed. They achieve in determining the reliability of each 3D joint position by employing a data fusion method based on Kalman filtering using multiple Kinect sensors. They take into account the measurement variance of noise for determining the contribution of an observation to the fused measurement. Additionally, they explain how to estimate the measurement variance for each one of the measurements. Finally, they present the average 3D position error of ten activities produced by their method, by a single Kinect and the average derived by multiple Kinect sensors, respectively. Almost in all cases, their method appears to give better results than the standard Kalman filter.
In [Berti et al (2014)] Kalman filtering is applied for robotic arms tracked by Kinect sensors. They denoise only the depth coordinate using Kalman filter methodology but they do not explain clearly the estimation process concerning the matrices involved. In their results presented through figures, it is obvious that the data are denoised, however the error reduction is not evaluated. In [Kong et al (2013)] a Kalman filter is described briefly to smooth 2D movements of a joint. In contrast with the aforementioned methods, the video data are derived via CCTV (Closed Circuit Television). Thus, they describe in which way they construct their joints' body model. Other scientists who are dealing with activity recognition via neural networks, use a simple Savitzky-Golay smoothing filter in order to correct the data [Yong et al (2015), Wentao et al (2016)]. This method is based on the previous, as well as the current and the two following observations.
In Section 2 the Kalman filter procedure along with the related likelihood function is provided. In Section 3 we describe the Tobit Kalman filter. In Section 4 the Kalman filter and Tobit Kalman filter approach for skeleton tracking is established. Finally, in Section 5, conclusions are presented.

2. KALMAN FILTER

In this section we describe briefly the Kalman filter (KF) [Peter (1979)], abbreviated as KF, which is used for estimating the unknown state vector $\mathbf{x} \in \Re^n$, of a discrete-time stochastic process that is governed by the linear stochastic difference equation

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

with a measurement (observation) $\mathbf{y} \in \Re^m$ given by

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

where $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q})$, $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R})$, and $N(\mu, \mathbf{\Sigma})$ denotes the normal distribution with mean value μ and covariance matrix $\mathbf{\Sigma}$. The matrices $\mathbf{A}, \mathbf{H}$ are the transition and observation matrices, respectively.
We define by $\hat{\mathbf{x}}_k^-$ the a priori state estimate at step k by assuming knowledge of the process history prior to step k and $\hat{\mathbf{x}}_k$ the a posteriori state estimate at step k by assuming that the measurement $\mathbf{y}_k$ is given.
The KF uses a form of feedback control; the filter firstly estimates the process state at some time and secondly it obtains feedback in the form of (noisy) measurements. So, the process of KF evolves in two stages: the predict stage and the update stage, determined by the associated equations:
The Predict Stage

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1}, \tag{3}$$

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q}. \tag{4}$$

where $\mathbf{P}_k^-$ and $\mathbf{P}_{k-1}$ are the covariance matrices of the errors of the a priori and a posteriori state estimates, respectively.
The Update Stage

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H}\mathbf{P}_k^- \mathbf{H}^T + \mathbf{R})^{-1}, \tag{5}$$

where $\mathbf{K}_k$ stands for the Kalman Gain (matrix), and

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_k^-), \tag{6}$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_k^-. \tag{7}$$

The error of the estimation for one step ahead and its variance are given by

$$\mathbf{u}_k = \mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_k^- and\ \mathbf{F}_k = \mathbf{H}\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}.$$

In applications of the filter, the measurement noise covariance $\mathbf{R}$ is usually measured prior to the filter operation or it is known. The determination of the process noise covariance matrix $\mathbf{Q}$ is more difficult and has to be estimated. In order to estimate $\mathbf{Q}$, the Maximum Likelihood Estimation (MLE) method can be used; the associated log-likelihood function for n measurements has the form [Tusell (2011), Proietti et al (2012)]

$$LogL(\mathbf{y}_1, ..., \mathbf{y}_n) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{k=1}^{n}(\log(|\mathbf{F}_k|) + \mathbf{u}_k^T\mathbf{F}_k^{-1}\mathbf{u}_k), \tag{8}$$

where n denotes the number of measurements and $|\mathbf{F}_k|$ the determinant of the matrix $\mathbf{F}_k$. The maximum likelihood estimators (MLEs) are most attractive because of their asymptotic properties. Under regularity conditions [Green (2002)], the maximum likelihood estimator has the following asymptotic properties:

- *Consistency*: the estimator $\hat{\theta}$ tends to a parameter θ_0 .
- *Asymptotic normality*: that is $\hat{\theta} \sim N(\theta_0, \mathbf{I}(\theta_0)^{-1})$, where

$$\mathbf{I}(\theta_0) = -E_0(\partial^2 lnL/\partial\theta_0^2)$$

- *Asymptotic efficiency*: $\hat{\theta}$ is asymptotically efficient and achieves the Cramer Rao lower bound.
- *Invariance*: The maximum likelihood estimator of $\gamma_0 = c(\theta_0)$ is $c(\hat{\theta})$ if $c(\theta_0)$ is a continuous and continuously differentiable function.

3. TOBIT KALMAN FILTER

In this section we describe the Tobit Kalman filter [Bethany (2014)], abbreviated as TKF, which provides a classification scheme for censored models [Bethany (2014), Tobin J (1958)]; these classes depend on the type of censoring and include also the cases of censoring, that depends on other variables. In the applications cases of censoring, the censored measurement model provides a measurement, either in knowing the exact value (it belongs to the uncensored region), or in knowing that the value lies into an interval.

In the general case of scalar measurements, the Tobit model is referred to as the censored regression model determined by (9),

$$y_k^* = hx_k + v_k,$$

$$y_k = \begin{cases} y_k^*, & T_{min} < y_k^* < T_{max} \\ T_{min}, & y_k^* < T_{min} \\ T_{max}, & y_k^* > T_{max}, \end{cases} \tag{9}$$

where y_k^* is the latent (hidden) variable, y_k is the measurement, h is an arbitrary scalar, T_{min}, T_{max} are the lower and upper thresholds-limits respectively and v_k is a Gaussian random variable with mean 0 and variance σ_v^2 . By (9) it is obvious that the Tobit Kalman Filter determines a non-linear process.

The standard KF doesn't provide optimal estimates of the unknown states vector when the measurements are censored; this happens because the assumptions of the KF are not met when the measurement is censored. In order to face the problem of censored measurements, we propose the TKF defined by [Bethany (2014)]

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

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

with

$$y_{k,i} = \begin{cases} y_{k,i}^*, & T_{min,i} < y_{k,i}^* < T_{max,i} \\ T_{min,i} & y_{k,i}^* < T_{min,i} \\ T_{max,i}, & y_{k,i}^* > T_{max,i}. \end{cases} \quad i = 1, 2, ..., m \tag{10}$$

where the noises $\mathbf{w}_k$ and $\mathbf{v}_k$ are defined by equations (1), (2) and $\mathbf{y}_k = (y_{k,i})_{i=1,...,m}$, $\mathbf{y}_k^*$ are defined as the saturated observation and latent variable respectively. Next we prove analytically for the case $\mathbf{H} = diag(h_1, ..., h_m)$ and $\mathbf{R} = diag(r_1^2, ..., r_m^2)$ the following useful Lemma 3.1.

Lemma 3..1. *The probability function of the i^{th} component of the measurement given the state vector is*

$$
\begin{aligned}
f(y_{k,i}|x_{k,i}) = \frac{1}{r_i}\phi\left(\frac{y_{k,i} - h_i x_{k,i}}{r_i}\right) u(y_{k,i} - T_{min,i}) u(T_{max,i} - y_{k,i}) \\
+\Phi\left(\frac{T_{min,i} - h_i x_{k,i}}{r_i}\right)\delta(T_{min,i} - y_{k,i}) \\
+\left(1 - \Phi\left(\frac{T_{max,i} - h_i x_{k,i}}{r_i}\right)\right)\delta(T_{max,i} - y_{k,i}),
\end{aligned}
$$

where ϕ and Φ are the probability and cumulative distribution function of standard normal distribution respectively, δ stands for the Kronecker delta function and u for the Heavyside function.

Proof. When the i^{th} component of the latent variable belongs to the uncensored region, (T_{min}, T_{max}), we get by (10) that

$$y_{k,i} = h_i x_{k,i} + v_{k,i}, \tag{11}$$

where $v_{k,i} \sim N(0, r_i^2)$. Thus by (11), the cumulative distribution for the measurement $Y_{k,i}$ is

$$
\begin{aligned}
F(y_{k,i}|x_{k,i}) &= P(Y_{k,i} \leq y_{k,i}), \\
&= P(h_i x_{k,i} + v_{k,i} \leq y_{k,i}) \\
&= P\left(\frac{v_{k,i}}{r_i} \leq \frac{y_{k,i} - h_i x_{k,i}}{r_i}\right) \\
&= \Phi\left(\frac{y_{k,i} - h_i x_{k,i}}{r_i}\right) \Rightarrow^{\partial y_{k,i}} \\
f(y_{k,i}|x_{k,i}) &= \frac{1}{r_i}\phi\left(\frac{y_{k,i} - h_i x_{k,i}}{r_i}\right).
\end{aligned} \tag{12}
$$

The probability of the i^{th} component of the latent variable to belong into the censored region from below is

$$
\begin{aligned}
P(y_{k,i} = T_{min,i}|x_{k,i}) &= P(y_{k,i}^* \leq T_{min,i}|x_{k,i}) \\
&= P(h_i x_{k,i} + v_{k,i} \leq T_{min,i})
\end{aligned}
$$

$$= P\left(\frac{v_{k,i}}{r_i} \leq \frac{T_{min,i} - h_i x_{k,i}}{r_i}\right)$$

and thus,

$$P(y_{k,i} = T_{min,i}|x_{k,i}) = \Phi\left(\frac{T_{min,i} - h_i x_{k,i}}{r_i}\right). \tag{13}$$

In the same way it is proved that

$$P(y_{k,i} = T_{max,i}|x_{k,i}) = 1 - \Phi\left(\frac{T_{max,i} - h_i x_{k,i}}{r_i}\right). \tag{14}$$

By (12)-(14) the probability distribution function for the measurements $\mathbf{y}_k$ is derived given the state vector $\mathbf{x}_k$. $\square$

We denote by $\mathbf{P}_{un,k}$, $\mathbf{P}_{min,k}$, $\mathbf{P}_{max,k}$ the probabilities of a measurement to be uncensored, or censored from below or censored from above, respectively, at time k. Then by *Lemma* 3.1 it is derived that

$$\mathbf{P}_{un,k} = diag \begin{bmatrix} \Phi\left(\frac{T_{max,1} - h_1 \hat{x}_{k,1}^-}{r_1}\right) - \Phi\left(\frac{T_{min,1} - h_1 \hat{x}_{k,1}^-}{r_1}\right) \\ \dots \\ \Phi\left(\frac{T_{max,m} - h_m \hat{x}_{k,m}^-}{r_m}\right) - \Phi\left(\frac{T_{min,m} - h_m \hat{x}_{k,m}^-}{r_m}\right) \end{bmatrix}, \tag{15}$$

$$\mathbf{P}_{min,k} = diag \begin{bmatrix} \Phi\left(\frac{T_{min,1} - h_1 \hat{x}_{k,1}^-}{r_1}\right) \\ \dots \\ \Phi\left(\frac{T_{min,m} - h_m \hat{x}_{k,m}^-}{r_m}\right) \end{bmatrix}, \tag{16}$$

$$\mathbf{P}_{max,k} = diag \begin{bmatrix} 1 - \Phi\left(\frac{T_{max,1} - h_1 \hat{x}_{k,1}^-}{r_1}\right) \\ \dots \\ 1 - \Phi\left(\frac{T_{max,m} - h_m \hat{x}_{k,m}^-}{r_m}\right) \end{bmatrix}. \tag{17}$$

By taking into account the matrices (15)-(17) and the properties of truncated normal distribution [Burkardt (2014), Tobin (1958)], we get that the expected value of the measurement when censored and uncensored measurements are included given the a priori estimation of the state vector has the form

$$\mathbf{E}(\mathbf{y}_k) = \mathbf{P}_{un,k}(\mathbf{H}\hat{\mathbf{x}}_k^- + \mathbf{R}^{\frac{1}{2}} l_k) + \mathbf{P}_{min,k}\mathbf{T}_{min} + \mathbf{P}_{max,k}\mathbf{T}_{max} \tag{18}$$

where $\mathbf{T}_{max} = (T_{max,i})_{i=1,..,m}, \mathbf{T}_{min} = (T_{min,i})_{i=1,..,m}$ and the parameter l_k at time k is the inverse Mill ratio [Burkardt (2014)],

$$l_k = \mathbf{P}_{un,k}^{-1} \begin{bmatrix} \phi\left(\frac{T_{max,1}-h_1\hat{x}_{k,1}^-}{r_1}\right) - \phi\left(\frac{T_{min,1}-h_1\hat{x}_{k,1}^-}{r_1}\right) \\ \dots \\ \phi\left(\frac{T_{max,m}-h_m\hat{x}_{k,m}^-}{r_m}\right) - \phi\left(\frac{T_{min,m}-h_m\hat{x}_{k,m}^-}{r_m}\right) \end{bmatrix}.$$

The covariance matrix of the measurement is given by

$$\mathbf{R}^* = \mathbf{R}\Big(\mathbf{I} + \mathbf{P}_{un}^{-1} diag(c_k) - diag(l_k)^2\Big) \tag{19}$$

where the parameter c_k [Burkardt (2014)] is given by

$$c_k = diag \begin{bmatrix} \frac{T_{min,1}-h_1\hat{x}_{k,1}^-}{r_1}\phi\left(\frac{T_{min,1}-h_1\hat{x}_{k,1}^-}{r_1}\right) - \frac{T_{max,1}-h_1\hat{x}_{k,1}^-}{r_1}\phi\left(\frac{T_{max,1}-h_1\hat{x}_{k,1}^-}{r_1}\right) \\ \dots \\ \frac{T_{min,m}-h_m\hat{x}_{k,m}^-}{r_m}\phi\left(\frac{T_{min,m}-h_m\hat{x}_{k,m}^-}{r_m}\right) - \frac{T_{max,m}-h_m\hat{x}_{k,m}^-}{r_m}\phi\left(\frac{T_{max,m}-h_m\hat{x}_{k,m}^-}{r_m}\right) \end{bmatrix}.$$

The Tobit Kalman Filtering process is defined as [Bethany A (2014)] :
The Predict Stage:

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1}, \tag{20}$$

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q}. \tag{21}$$

The Update Stage:

$$\mathbf{R}_1 = \mathbf{P}_k^-\mathbf{H}^T\mathbf{P}_{un,k},$$

$$\mathbf{R}_2 = \mathbf{P}_{un,k}\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T\mathbf{P}_{un,k} + \mathbf{R}_k^*,$$

$$\mathbf{K}_k = \mathbf{R}_1\mathbf{R}_2^{-1}, \tag{22}$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{E}(\mathbf{y}_k|\hat{\mathbf{x}}_k^-)), \tag{23}$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{P}_{un,k}\mathbf{H})\mathbf{P}_k^-. \tag{24}$$

4. IMPLEMENTATIONS

In the present paper we use the Microsoft Kinect sensor 2 to record 3D point sequences of a human skeleton in motion and our aim is to denoise the coordinates for every joint in order to improve the representation of the movements. For this reason, in our first approach, we use a KF for each one of the joints separately; the input includes the joints' coordinates $[x, y, z]$ (measurement) and the outputs the denoised coordinates (state vectors).
Thus, we define the transition matrix $\mathbf{A}$ and the observation matrix of the model, $\mathbf{H}$ as

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

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Next, we have to estimate the covariance matrix for the process noise, $\mathbf{Q}$. Firstly we assume that the entries of the covariance matrix of the measurement noise, $\mathbf{R}$, are of the order 10^{-2}. Then, by the likelihood function (8), the entries of the matrix $\mathbf{Q}$ can be derived. Interestingly we notice by various joints' movements, that the entries of $\mathbf{Q}$ appear to be smaller than those of matrix $\mathbf{R}$, and generally they depend on the accuracy of the Kinect sensor and the joints' speed. Concerning slow motions, the values are experimentally found to be smaller than 10^{-4} and for faster motions they lie between 10^{-3} and 10^{-2}. In order to create a general model for denoising Kinect's measurements, in which we will not estimate the matrix $\mathbf{Q}$ for every time-window, because this is time consuming, we can assume that

$$\mathbf{Q} = 0.002 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In our first experiment, a man throws a ball with his right hand, and this movement is recorded by Kinect; the video consists of 266 frames (almost 8.8667 sec.). Many joints and especially the joints on the right side were self-occluded. So, we used the KF as described in Section 2 in order to denoise the data, i.e., to reveal the hidden coordinates due to the occlusion. It is obvious by $Fig.$1 and 2 that the KF smooths the spatial coordinates without affecting the movement, i.e., it does not provide oversmoothing of the motion.

In other experiments, an individual seats in front of the Kinect sensor and the skeleton appears to fall down unnaturally; apparently this is due to some noise of Kinect. So, we used Kalman filtering, but the noise could not be corrected satisfactorily. In order to correct the noise, we studied many recordings by the groundtruth sensor, Vicon; we observed that the velocity of spatial coordinates x and z did not exceed 31 cm per two consecutive frames for every joint, while the coordinate y did not exceed 18 cm respectively. Thus we took these restrictions into account, in order to correct the data. So we constructed a TKF with limits $\mathbf{T}_{min}$ and $\mathbf{T}_{max}$ for the spatial coordinates $[x, y, z]$ as follows,

$$\mathbf{T}_{max,k} = (\hat{x}_{k-1} + 0.31, \hat{y}_{k-1} + 0.18, \hat{z}_{k-1} + 0.31),$$

$$\mathbf{T}_{min,k} = (\hat{x}_{k-1} - 0.31, \hat{y}_{k-1} - 0.18, \hat{z}_{k-1} - 0.31),$$

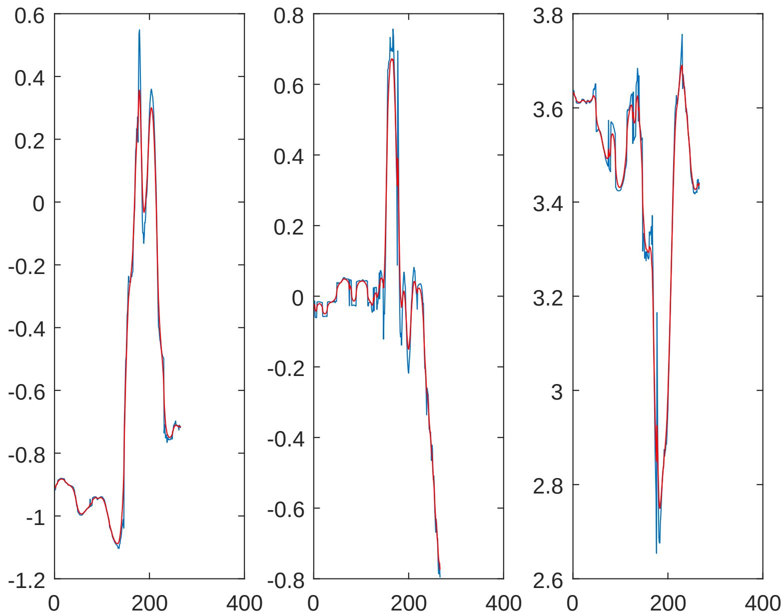

Figure 1: *The spatial coordinates for the right hand as they result by Kinect (blue) and the Kalman filter (red)*

where $\mathbf{T}_{max,k}$ and $\mathbf{T}_{min,k}$ are the limits of the TKF at time k which depend on the previous estimation of spatial coordinates. So by (10), for the measurement $\mathbf{y}_k = [x_k, y_k, z_k]$ at time k we get

$$y_{k,i} = \begin{cases} y^*_{k,i}, & T^i_{min,k} < y^*_{k,i} < T^i_{max,k} \\ T^i_{min,k}, & y^*_{k,i} < T^i_{min,k} \\ T^i_{max,k}, & y^*_{k,i} > T^i_{max,k}. \end{cases} \quad i = 1, 2, 3$$

The aforementioned TKF model can appropriately smooth big abnormal movements due to Kinect's errors. Apparently, if $T^i_{min,k} \to -\infty$ and $T^i_{max,k} \to \infty$ (i.e., the range of TKF's state values becomes too big) the TKF becomes the standard KF. As can be seen in $Fig.3$ the skeleton motion of the TKF (green) does not exhibit any unexplainable fall. Notice that the standard KF can correct the noise but not as well as the TKF. In our experiments, the TKF exhibits skeleton falls till almost 4 cm, which is an acceptable bound. On the other hand the standard KF exhibits skeleton falls more than 4 and till 8 cm per (two) consecutive frames, which is not realistic. This conclusion is more clear in $Fig.4$, where the head's spatial coordinate y_k for each frame k is illustrated.

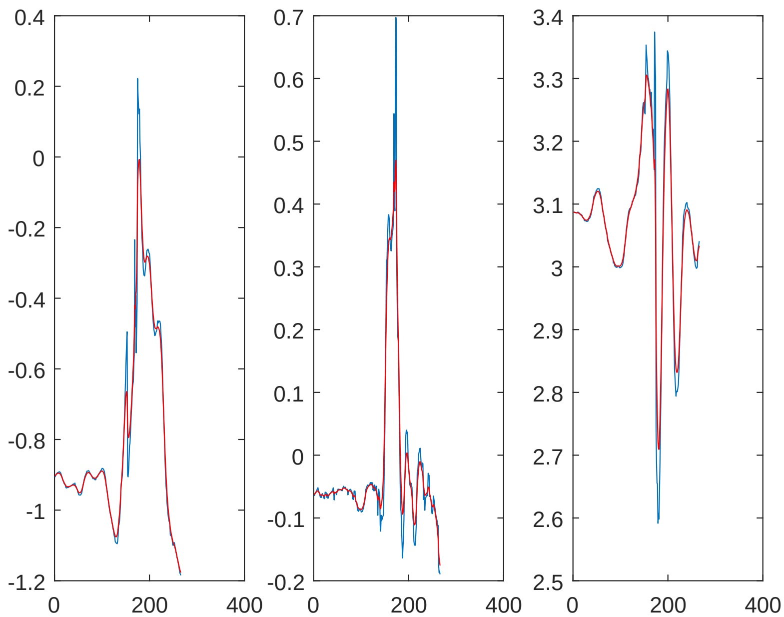

Figure 2: *The spatial coordinates for the left hand as they result by Kinect (blue) and the Kalman filter (red)*

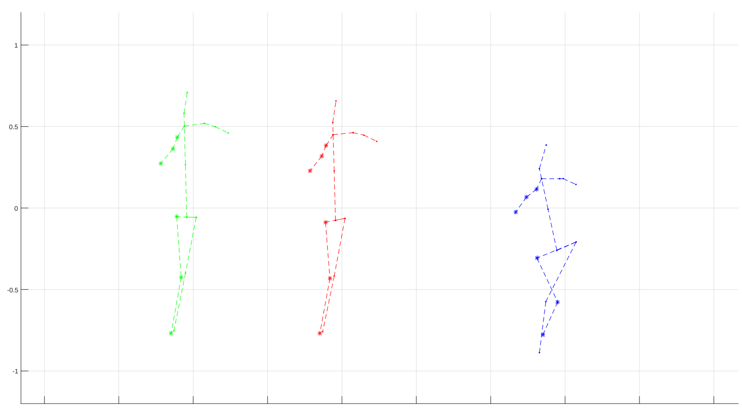

Figure 3: *The skeletons of Tobit Kalman filter(green), Kalman filter (red) and Kinect(blue) at the frame of the fall.*

5. CONCLUSION

The aim of this paper is to improve skeleton tracking, using a single Kinect sensor, which generates error in recordings due to occlusion, self-occlusion e.t.c.. So, we propose to use a Tobit Kalman filter for skeleton tracking in real time.

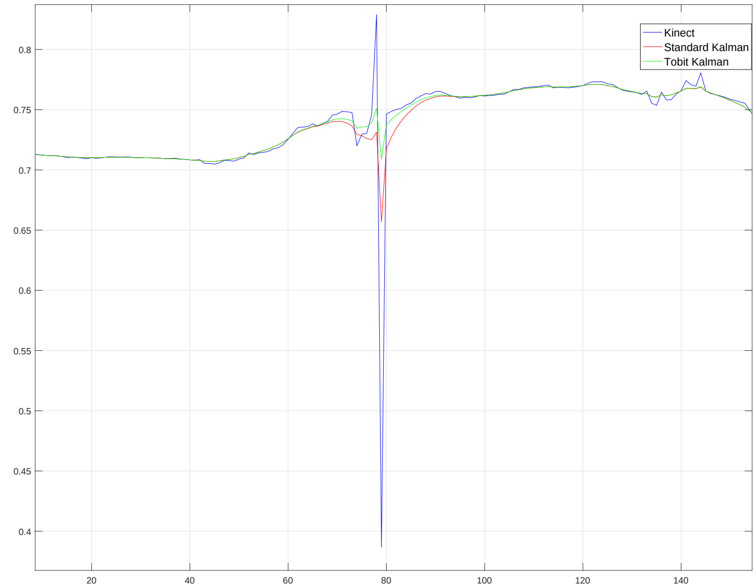

Figure 4: *The head's spatial coordinate y_k for each frame k of Tobit Kalman filter (green), Kalman filter (red) and Kinect (blue).*

In this approach we have to define the limits $\mathbf{T}_{max,k}$ and $\mathbf{T}_{min,k}$ in a reasonable manner for evrey time k. For that purpose we considered a lot of skeleton data, with various joints' movements, which were obtained by means of the groundtruth sensor, Vicon.
The covariance matrix of the noise process $\mathbf{Q}$, using Kalman filtering procedure was estimated via maximum likelihood estimation. Between the two filters, i.e., the standard Kalman filter and the Tobit Kalman filter, the latter was more accurate performing a better skeleton tracking. Furthermore, in some frames when the skeleton collapsed due to occlusion, the method of the Tobit Kalman filter proposed, corrected better the error in recordings than the standard Kalman filter.

ΠΕΡΙΛΗΨΗ

Στην παρούσα εργασία, αναλύουμε δεδομένα από την κάμερα Microsoft Kinect 2 χρησιμοποιώντας φίλτρα Kalman και Tobit Kalman για την ελαχιστοποίηση του θορύβου που εμφανίζεται στα δεδομένα. Τα δεδομένα αφορούν τριδιάστατες χωρικές συντεταγμένες που καταγράφουν κινήσεις των αρθρώσεων ενός ανθρώπου, στις οποίες εμφανίζονται σφάλματα στην ακρίβεια των μετρήσεων. Χρησιμοποιούμε πέ-

ρα από το κλασικό φίλτρο Kalman και το Φίλτρο Tobit Kalman, προκειμένου να συμπεριλάβουμε στο μοντέλο περιοριστικές συνθήκες με βάση τα ανθρωπομετρικά στοιχεία, και συγκεκριμένα τις αποστάσεις μεταξύ διαφόρων αρθρώσεων. Στο τέλος παρουσιάζουμε προσομειώσεις για την κίνηση του σκελετού πριν και μετά τη χρήση των φίλτρων.

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

REFERENCES

Berti, E. M., Salmeron, A. J. and Benimeli, F. (2012). Kalman Filter for Tracking Robotic Arms Using low cost 3D Vision Systems. *The Fifth International Conference on Advances in Computer-Human Interactions.* 236-240.

Bethany, A. (2014). *Tobit Filter Kalman,* University of Delaware: Department of Electrical and Computer Engineering.

Burkardt, J. (2014). *The Truncated Normal Distribution,* Florida State University: Department of Scientific Computing.

Galna, B., Barry, G., Jackson, D., et. al. (2014.) Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson'As disease. *Gait and Posture* Vol 39. Issue 4. 1062-1068.

Green, W. H. (2002). *Econometric Analysis,* 5th edition, Chapter 17, New York: Prentice Hall.

Kong, W. A., Hussain H. and Md Saad, M. H.(2013.) Essential Human Body Points Tracking Using Kalman Filter. *Proceedings of the World Congress on Engineering and Computer Science* Vol I.

Peter, S. (1979.) Stochastic models, estimation, and control. *Mathematics in Science and Engineering* Vol 141.

Proietti, T. and Luati, A. (2012). Maximum Likelihood Estimation of Time Series Models: The Kalman Filter and Beyond. *Munich Personal RePEc Archive* No 41981.

Sungphill, M., Youngbin, P., Dong, W. K., et. al. (2016). Multiple Kinect Sensor Fusion for Human Skeleton Tracking Using Kalman Filtering. *International Journal of Advanced Robotic Systems.*

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society* 24-36.

Tusell, F. (2011). Kalman Filtering in R. *Journal of Statistical Software* Vol 39. Issue 2.

Wentao, Z., Cuiling, L., Junliang, X., et. al. (2016). Co-occurrence Feature Learning for Skeleton based Action Recognition using Regularized Deep LSTM Networks. *The 30th AAAI Conference on Artificial Intelligence.*

Yong, W. D. and Wei, W. (2015). Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition. *Conference on Computer Vision and Pattern Recognition* 1110-1118.

Zhang, C. and Zhang, L. (2013). Activity Recognition in Smart Homes Based on Second-Order Hidden Markov Model. *International Journal of Smart Home* Vol 7, No 6, 237-244.

Zhen, L., Zhiqiang, W., Yaofeng, Y. W. et al. (2015). An Adaptive Hidden Markov Model for Activity Recognition Based on a Wearable Multi-Sensor Device. *Journal of Medical Systems-Springer*,39-57.