# Automatic Noise Estimation and Context-Enhanced Data Fusion of IMU and Kinect for Human Motion Measurement

Ali Akbari, Xien Thomas, Roozbeh Jafari

Embedded Signal Processing Lab, Texas A&M University, TX

{aliakbari, xien.thomas, rjafari}@tamu.edu

***Abstract*— The aim of this paper is to propose a robust, accurate and portable system for human body motion measurement. The system includes Inertial Measurement Units (IMU) and a Kinect or vision sensors. Since Kinect sampling rate is low (30 Hz per second) and it suffers from occlusion, it cannot individually measure human body motion accurately and robustly. On the other hand, IMU does not suffer from these problems, but it suffers from drift in particular with long-term motion monitoring and other types of errors (*e.g.,* high acceleration motions, temperature and voltage variations). Thus, in this study, IMU and Kinect data were fused using a context enhanced extended Kalman filter. Rules were generated based on the context of motion in order to adjust Kalman filter parameters. In addition, an automated approach is introduced to estimate the variance of the noise of the sensors during the operation. Considering motion context and automatic noise detection, the robustness of monitoring is enhanced against errors related to motion context (*i.e.,* high acceleration and long-term motions); furthermore, offline calibration is no longer required to set the parameters of the filter. The system was tested on leg and arm motions. The root mean square error of our fusion method was $6.08^{\circ}$ lower than using only gyroscope, $16.98^{\circ}$ lower than using only accelerometer, $2.49^{\circ}$ lower than using only the Kinect and $8.99^{\circ}$ lower than using simple EKF fusion method, which does not consider motion context and automatic noise estimation.**

## I. Introduction

Human body motion measurement is of principal importance due to the wide number of applications in health monitoring, human motion research, physiotherapy, activity detection, sport, robotics, virtual reality and the animation industry. Different approaches have been used to measure human motions and limb angles including acoustic, inertial, textile, magnetic, mechanical, optical and radio frequency systems [1].

In the field of human body motion measurement for biomedical applications and kinematics research, motion cameras have been widely used. This type of motion measurement is very accurate, however, it is limited to a laboratory environment [2]. Furthermore, this approach needs some preparation and manual adjustments, so it is very time consuming. These limitations, in addition to the high costs of motion camera systems, has promoted the use of portable and low cost systems to measure motion over a longer period of time and outside of laboratory environments. These systems can be used for daily human activity monitoring.

Among portable systems, inertial sensors (*i.e.,* accelerometers and gyroscopes) are very popular [1]. These sensors are low cost, lightweight and portable, but they suffer from a high level of noise and errors. Two main sources of errors are due to drift, as a result of integration, and high acceleration motions. Integration at the output of the gyroscope causes accumulation of the bias. Hence, the angle calculated from the gyroscope becomes completely unusable in a short while. On the other hand, accelerometer uses the Earth's gravity to measure the angle. It does not generate drift, however any acceleration created by human motions produce large errors in the angle calculated from the accelerometer, as the sensor will be report the combination of gravity and acceleration due to movements.

To overcome these issues, researchers have used complementary features of these sensors by fusing gyroscope and accelerometer data. Furthermore, some researchers have added other types of sensors such as magnetometers, GPS and vision based systems like the Kinect in order to improve the performance. Among all complementary filters in this field, the Kalman filter has been used extensively by researchers. [3-8]. Nevertheless, two issues have not yet been adequately addressed in prior investigations. One is the realistic evaluation of variable and uncertain sensor noises under different conditions. The other is the negative impact of the motion context on sensor performance.

Properly measuring and modeling the noise of the sensors is very important when using the Kalman filter. Realistically modelling the sensor error is very difficult due to the variable sensor performance in various contexts. Thus the errors resulting from mismodeling, non-modeling and non-white properties of input data are significant as shown in prior investigations [9]. Traditional solutions for noise estimation have some practical limitations since they require laboratory setup and a calibration process. Moreover, those approaches do not consider variations in the noise under various conditions.

Equally important, the condition of the motion has a large impact on the measurement accuracy. For instance, high accelerations in movements cause large error in calculation of the angle extracted from the accelerometer. Also, longer motions allow accumulation of huge drift in the angle coming from the gyroscope. Moreover, when occlusion occurs, the accuracy of the Kinect degrades drastically. We refer to these

978-1-5090-6244-7/17/$31.00 ©2017 IEEE 

motion conditions as context. Any variable, situation or condition that may affect or constrain the outcome is known as context [10].

The correct operation of the monitoring systems under various conditions is of principal importance. For example, there are rehabilitation routines to restore the strength of patient's back and to help them gradually regain their physical abilities. Orthopaedic surgeons and physical therapists may recommend that the patient perform the exercises for 10 to 30 minutes a day, and one to three times a day during early recovery. It is important to monitor patient's motions for the duration of the exercise in order to detect the recovery status and provide feedback when necessary. On the other hand, in sport research, for example in golf, the motions are very quick and include high accelerations. The sensors' accuracy degrades in these conditions. So it is essential to consider the impact of the motion context on performance of the sensors.

This paper aims to deal with aforementioned issues fusing IMU and Kinect data for human body motion measurement. An adaptive extended Kalman filter (EKF) was designed to fuse the sensor data. The proposed system is able to detect the context, and set required EKF parameters based on the motion context. To accomplish this, it is required to determine the variance of the noise appropriately for the EKF. Therefore, an automated and online technique is proposed to estimate the noise of the sensors while the system is running. Thus, the offline calibration process can be eliminated, additionally, the system can also take into consideration the noise variations due to environment changes as the monitoring is in progress.

The principal contributions of this paper are:

- A set of rules are proposed to detect the context of motion. The variance of the noise required for EKF are determined based on the motion context.
- The variance level of the noise is estimated while the system is running. To clarify, as the variance of the noise of each sensor changes in different environments and motions, a set of rules to track the variance of the noise is selected without using external equipment while the system is running.
- The accuracy of the operation and our proposed techniques is evaluated in various conditions which usually degrade sensor performance, *e.g.* in presence of high acceleration, fast and long-term motions. These types of motion-based errors have not been considered in details in prior investigations. A set of tests were performed on human limbs to assess system performance under various conditions.

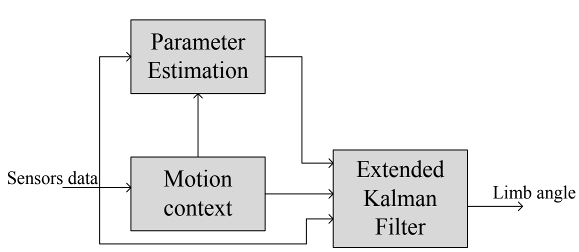


Figure 1 – Block diagram of the proposed system

## II. METHODS

The block diagram of the entire system is shown in Figure 1. The variances of the noise are estimated and updated in different conditions while the system is running. This information is retained and used by the EKF whenever system observes a similar context at a later time. At each time step, the system first detects motion context. It then uses the estimated variance of the noise which fits that condition in the EKF to calculate the body limbs angle.

### *A. Motion context*

The first stage of this operation is the motion context detection. This process fuses sensor data in order to detect the context including: fast, slow and normal motions, stationary, high acceleration motions, long-term motions, and Kinect self-occlusion. Herein, a set of rules are designed to identify various contexts. Afterward, additional rules are introduced to set the variance of the noise of the sensors based on the motion context.

The speed of the motion changes gradually which prohibits to clearly distinguish the boundaries between slow, normal and fast motions. Therefore, a fuzzy system is utilized for speed level detection. This system converts gyroscope, accelerometer and Kinect data to a fuzzy set using fuzzy linguistic variables including *{stationary, slow, normal, fast}*. For the purpose of computational simplicity, triangle membership functions are used in the fuzzy system [9]. Furthermore, a simpler fuzzy system is used for high acceleration detection. Kinect and accelerometer data are used as inputs and the fuzzy set is {high acceleration, not high acceleration}.

With longer motions (*i.e.,* motions lasting for a few minutes), the drift of the gyroscope grows rapidly. When the subject is stationary the EKF does not use the gyroscope, and therefore, the operation does not suffer from the drift accumulation or error. However, when the subject begins the motion, the gyroscope drift integration error is observed. In this case, a timer measures the duration of the motion. If the duration of the motion is larger than a threshold, the system identifies it as a long-term motion.

Self-occlusion occurs when the view of a body segment is obstructed by other body segments. When the occlusion occurs, Kinect attempts to estimate the location of the covered joint which would increase the error. Self-occlusion is primarily due to the Kinect's sensitivity to the subject's body rotation with respect to the camera [11]. To detect this situation, the angle between the two shoulders is measured and compared to a threshold Equation 1.

$$\alpha = \tan^{-1}\left(\frac{Z_{\mathrm{shR}} - Z_{\mathrm{shL}}}{X_{\mathrm{shR}} - X_{\mathrm{shL}}}\right) \quad (1)$$

Where $Z_{shR}$ is the $z$ component of the right shoulder's position and $X_{shL}$ is the $x$ component of the left shoulder's position in the Kinect's coordinate system.

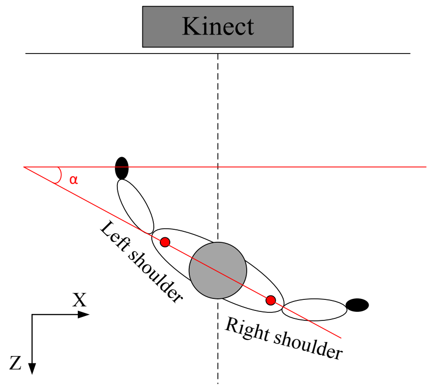


Figure 2 – The angle between the two shoulders when occlusion occurs

The angle between the two shoulders is shown in Figure 2. It has been shown that as this angle increases, the Kinect error increases and when angle is greater than 50° , the measurements becomes unstable and unreliable [11]. Therefore, our threshold for self-occlusion detection is 50° .

The first utility of the motion context detection would be tuning the gain of EKF and adjusting the variance of the noise properly. The second utility is leveraging the motion context to estimate the variance of the noise while the system is in operation.

*B. Parameter estimation*

Following the detection of the motion context, the EKF sets variance of the noise for sensors corresponding to the determined context. The parameter estimation block attempts to estimate the real value on the variances for each condition. The EKF can reuse these values whenever it observes similar condition at a later time.

Typically for every context, one sensor is expected to be more accurate and stable compared to the others. For every specific context, the accurate sensor is used to estimate the noise in other sensors. This process uses certain predefined rules to estimate the noise variance of the sensors. The estimated values are saved in a look-up table. The look-up table is continuously updated while the system is running. The values are retained for the following runs. This table is utilized by the EKF block in order to set the covariance matrix of the process model and measurement model noise in every time step. As mentioned before, these variances are important to determine the level of contribution of each sensor at each time step. More precise noise estimation leads to more accurate tracking and in our case, more accurate angle calculation. The advantage of this approach is that when the sensor performance varies due to voltage or temperature variations, the system can update the values.

The rules for the online estimation of the variance on the sensor noise are as follows:

- In a stationary condition, the gyroscope's bias and all sensor typical noise variance are estimated. This typical variance will be used when the subject moves slowly or at a normal speed. The output of the sensors should be stable in this condition, as no movement is present; therefore, any variation is noise. Moreover, the gyroscope output should be zero, and consequently, any other value represents the bias.
- In a stationary condition, the noise variance for the gyroscope is set to infinity. Thus, the EKF does not use gyroscope integration in a stationary condition.
- When the subject begins the motion from a specific orientation (A) and stops at a specific orientation (B), the angle difference can be measured by the Kinect (or accelerometer when the Kinect is not available). It has been known that the Kinect/accelerometer is more reliable than the gyroscope in stationary conditions [12]. Hence, the angle difference between orientation (A) and (B) is measured by gyroscope integration and compared to the Kinect/accelerometer angle measurement to update the noise variance for the gyroscope. This rule can determine the noise variance for the gyroscope in presence of slow, normal or fast motions.
- When the subject is moving and no high acceleration movement is present, the angle is extracted from the accelerometer and the Kinect. Then the derivatives of the angles, equivalent to the angular velocity, are compared to gyroscope data to estimate the variance of the noise of accelerometer and Kinect during this type of motion. This rule can determine the variances of noise of the accelerometer and Kinect in slow, normal and fast motions.
- The accelerometer noise variance in high acceleration state is set to infinity and hence the EKF ignores the accelerometer when motions with high acceleration are present.
- The Kinect is more accurate over longer time. Therefore, the mean angle from the Kinect is more reliable than the same measure acquired through the gyroscope. Gyroscope noise variance in long-term motions is estimated based on the difference between gyroscope mean angle and Kinect mean angle.
- When self-occlusion is detected, the variance of the noise of Kinect would be set to infinity.

*C. Extended Kalman Filter*

This process fuses data at the sensor level in order to estimate human limb angle. The information from the other blocks helps to adjust fusion filter noise parameters. In every time step, the EKF first reads the variance values from the look-up table updated by the parameter estimation block. It then sets the covariance matrix of process and measurement noise based on the current motion context. Finally, the EKF calculates and estimates the accurate limb angle from noisy data.

The state model in this study utilizes the quaternion representation of orientation and components of angular velocity.

$$X = \begin{bmatrix} q & \omega \end{bmatrix}^T \quad (2)$$

Where $q$ is a four-element vector of quaternions and $\omega$ is a three-element vector containing angular velocity (degree/sec) in three directions. The quaternion derivatives formula describes the quaternion process. Also, a simplified human motion model [6], which models angular accelerations as a first order random walk process, is used for angular velocity, as shown in Equations 3 and 4.

$$\dot{q} = \frac{1}{2} q \otimes [0 \quad \omega_x \quad \omega_y \quad \omega_z] \quad (3)$$

$$\dot{\omega}_{x,y,z} = -\frac{1}{\tau}\omega_{x,y,z} + v \quad (4)$$

where $\tau$ is time constant, describing how fast angular rates typically change for each axis. This constant depends on the type of application and could be heuristically customized for each application. $v$ is the driving noise.

The measurement model is same as the state model, with the only difference that there are $n$ quaternions equal to the number of motion sensors in the measurement model, as shown in Equation 5. Therefore, this model is scalable and can accommodate the scenarios where the number of sensors increase. Herein, we have two motion sensors including accelerometer and Kinect in our measurement model. However, any other orientation sensors such as magnetometer or vision based measurement systems can be added to this system.

$$Z = \begin{bmatrix} q_{s1} & q_{s2} & \dots & q_{sn} & \omega_g \end{bmatrix}^T \quad (5)$$

Where $\omega_g$ is the output of gyroscope (in degree/sec).

The covariance matrix of the noise for state and measurement models are the most important part of the EKF. The EKF gain is strongly correlated with the noise variance values. In the EKF, the ratio of the measurement and state model noise is important, and not the absolute values. The sensor with higher noise variance has less weight, and conversely, sensor with smaller variance gets a larger weight.

Herein, the variances are acquired from the look-up table created by the parameter estimation block and covariance matrices are formed.

## III. MEASUREMENT RESULTS

A Vicon motion camera system was used as the ground truth to validate the system performance. A set of tests and physical activities were performed by a healthy subject's involving movements in right leg and right arm.

Leg tests included right hip and right knee flexion and extension in range of motion while the subject was standing. The subject performed each test twice at different speeds separated by a break. In each test, the subject flexed and extended his hip 10 times with a short standstill between them. The subject also walked at two speed levels: normal and fast. Moreover, in one long-term test, the subject walked for 5 minutes with short pauses, and on occasions the subject pretended to kick a ball to create a motion with high acceleration. Two IMUs were placed on the shank and the thigh.

Arm tests included lower arm flexion and extension at two different speeds: slow and fast, flexion and extension with sudden start and stop to create motions with high acceleration, and a long-term (5-minute) flexion/extension routine with periodic pauses. Each test was repeated twice and every test included 10 flexion/extension motions. Since the upper arm was motionless during the arm test, just one sensor was placed on the lower arm.

The average root mean square error (RMSE) of flexion/extension angle leveraging our fusion technique showed improvement in comparison to the estimates from individual sensors. The average angle RMSE in degrees and the standard deviations are shown in Table I. In addition, the RMSE of our fusion method, which considers motion context and automated noise estimation, is compared to a simple fusion filter without leveraging context detection and noise estimation. This simple method contained EKF similar to our fusion technique; however, the covariance matrices for process model and measurement model noise remained constant.

Under slow and normal motions, the difference between our approach and simple fusion approach was not significant. However, with other motion scenarios, where the context of the motion impacted the sensor error, the accuracy of the simple fusion technique degraded significantly.

In slow and normal motions, the RMSE of the gyroscope was higher than other sensors. Our fusion method estimated the noise online and adjusted the Kalman gain to reduce the impact of the noise from gyroscopes. The accuracy of the fusion in this scenario is better than all individual sensors.

In fast motions, the accelerometer accuracy reduced drastically. The accuracy of the Kinect and gyroscope also reduced, but they were still more accurate than the accelerometer. In this case the fusion algorithm relied less on the accelerometer. However, the simple fusion method did not consider the context and it used noise variances similar to normal motions. As a result, the accuracy of simple fusion method was affected negatively.

Table I –Average RMSE error in angle (standard deviation)

| Test | Our fusion method | Gyro | Accelero -meter | Kinect | Simple fusion |
|---|---|---|---|---|---|
| **Slow and Normal motions** | 2.91 (1.05) | 5.01 (2.45) | 4.53 (1.81) | 4.06 (1.33) | 3.65 (1.50) |
| **Fast motions** | 7.56 (1.7) | 10.97 (3.93) | 17.67 (5.65) | 11.85 (4.13) | 9.36 (4.01) |
| **Long-term motions** | 5.57 (1.93) | 21.54 (14.9) | 6.22 (1.99) | 7.49 (2.02) | 6.13 (1.95) |
| **High acceleration motions** | 7.84 (2.27) | 10.71 (5.92) | 63.39 (28.59) | 10.42 (1.21) | 40.7 (2.71) |

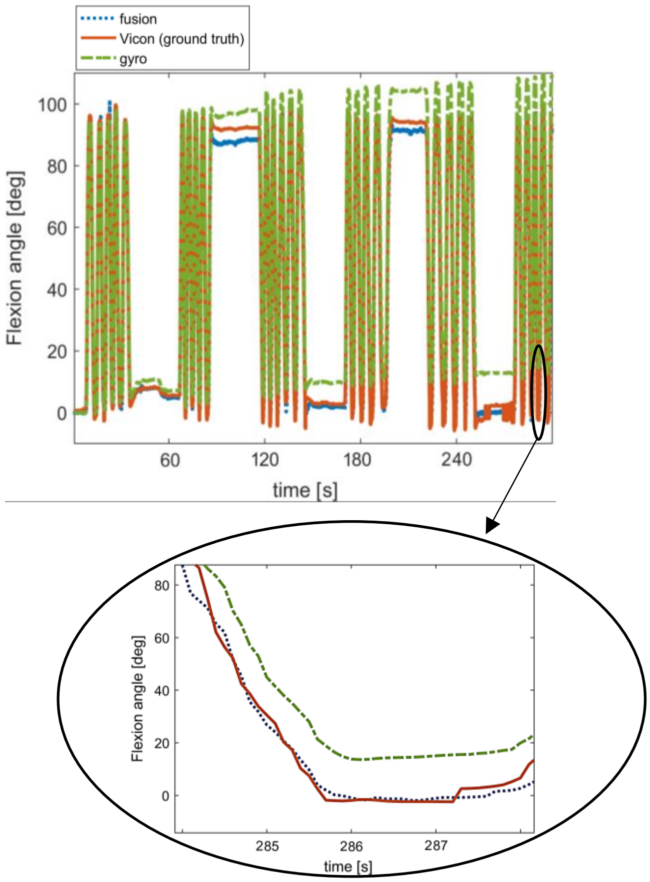


Figure 3 –long-term arm flexion and drift effect on gyroscope

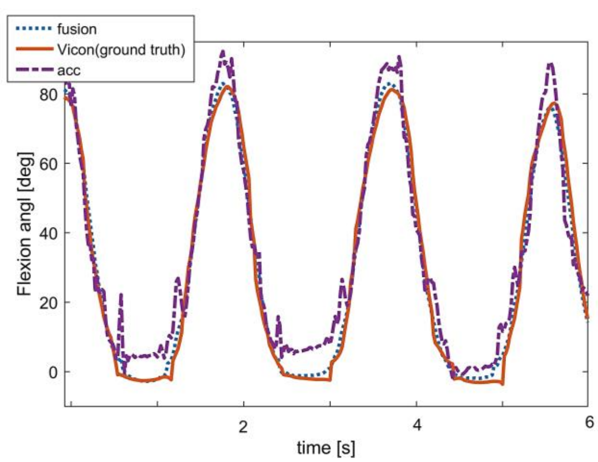


Figure 4 - Thigh high acceleration flexion and its effect on accelerometer. Accelerometer angle has big error in the beginning and end of movement (peaks and valleys)

In long-term motions, the gyroscope's angle error increased as time increased, while other sensors were more accurate, as shown in Figure 3. It can be seen that the fusion angle is closer to the ground truth (Vicon) and it does not track the angle of the gyroscope as time grows.

Finally, high acceleration motions caused very large error in accelerometer data for angle measurement. The proposed system was able to detect this context and correct the error by neglecting the accelerometer data. Figure 4 shows the high error of accelerometer. The fusion algorithm relied on the Kinect measurements rather than the other sensors and tracked the ground truth angle more closely. The simple fusion method used constant weight for the accelerometer, and therefore the accuracy was impacted negatively.

## IV. CONCLUSION

A context-aware data fusion method based on extended Kalman filter is presented in this paper to measure human body motion. This method considers the impact of motion context on sensor performance. In addition, an automatic online approach is proposed to estimate the noise variance that will be supplied to the EKF. As a result, the system does not require any adjustment or offline calibration prior to deployment. Additionally, in contrast to prior investigations that determine the parameters of the EKF by trial and error, or specific to particular applications, our method considers context changes and variation to estimate the noise variance on-the-fly. We validated our proposed techniques on human subjects and established its effectiveness and the added robustness that it offers.

## REFERENCES

[1] Bo, A., M. Hayashibe, and P. Poignet. Joint angle estimation in rehabilitation with inertial sensors and its integration with Kinect. in EMBC'11: 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011. IEEE.

[2] Mokhlespour Esfahani, M.I., et al., Novel printed body worn sensor for measuring the human movement orientation. Sensor Review, 2016. 36(3): p. 321-331.

[3] Marins, J.L., et al. An extended Kalman filter for quaternion-based orientation estimation using MARG sensors. in Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on. 2001. IEEE.

[4] Foxlin, E. Inertial head-tracker sensor fusion by a complementary separate-bias Kalman filter. in Virtual Reality Annual International Symposium, 1996., Proceedings of the IEEE 1996. 1996. IEEE.

[5] Luinge, H.J. and P.H. Veltink, Measuring orientation of human body segments using miniature gyroscopes and accelerometers. Medical and Biological Engineering and computing, 2005. 43(2): p. 273-282.

[6] Roetenberg, D., et al., Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation. IEEE Transactions on neural systems and rehabilitation engineering, 2005. 13(3): p. 395-405.

[7] Li, W. and J. Wang, Effective adaptive Kalman filter for MEMS-IMU/magnetometers integrated attitude and heading reference systems. Journal of Navigation, 2013. 66(01): p. 99-113.

[8] Mokhlespour Esfahani, M.I., et al., Trunk Motion System (TMS) Using Printed Body Worn Sensor (BWS) via Data Fusion Approach. Sensors, 2017. 17(1): p. 112.

[9] Wang, J. and Y. Gao, The aiding of MEMS INS/GPS integration using artificial intelligence for land vehicle navigation. IAENG International Journal of Computer Science, 2007. 33(1): p. 61-67.

[10] Snidaro, L., J. García, and J. Llinas, Context-based information fusion: a survey and discussion. Information Fusion, 2015. 25: p. 16-31.

[11] Glonek, G. and A. Wojciechowski. Kinect and IMU Sensors Imprecisions Compensation Method for Human Limbs Tracking. in International Conference on Computer Vision and Graphics. 2016. Springer.

[12] Plamondon, A., et al., Evaluation of a hybrid system for three-dimensional measurement of trunk posture in motion. Applied Ergonomics, 2007. 38(6): p. 697-712.