Motion model transitions in GPS-IMU sensor fusion for user tracking in augmented reality

Erkan Bostanci

Abstract—Finding the position of the user is an important processing step for augmented reality (AR) applications. This paper investigates the use of different motion models in order to choose the most suitable one, and eventually reduce the Kalman filter errors in sensor fusion for such applications where the accuracy of user tracking is crucial. A Deterministic Finite Automaton (DFA) was employed using the innovation parameters of the filter. Results show that the approach presented here reduces the filter error compared to a static model and prevents filter divergence. The approach was tested on a simple AR game in order to justify the accuracy and performance of the algorithm.

Keywords—Deterministic Finite Automaton (DFA), sensor fusion, GPS, IMU, model selection, augmented reality.

I. INTRODUCTION

INTEGRATION of data from Global Positioning System (GPS) and Inertial Measurement Unit (IMU) sensors has been well-studied [1]–[3] in order to improve upon the robustness of the individual sensors against a number of problems related to accuracy or drift. The Kalman filter (KF) [4] is the most widely-used filter due to its simplicity and computational efficiency [5] especially for real-time user tracking applications such as AR.

Attempts have been made to improve the accuracy of the filter using adaptive values for the state and measurement covariance matrices based on the innovation [6] and recently fuzzy logic was used for this task [7], [8]. In some studies [9], [10] used dynamic motion parameters to decide on the dominance of individual sensors for the final estimate.

Alternative approaches suggest using different motion models for recognizing the type of the motion [11]–[15]. Some of these studies (*e.g.* [12], [13] use a Bayesian framework for identifying a scoring scheme for selecting a motion model and some other studies, see [15], apply different motion models concurrently and select one of them according to a probabilistic approach.

This paper presents the selection and use of different motion models according to a DFA model [16] in order to reduce the filter error and ensure faster filter convergence. The rest of the paper is structured as follows: Section II presents the methods used for obtaining positional estimates from individual sensors. The fusion filter which uses these motion estimates in order to produce a single output is presented in Section III. Results are given in Section V and an AR application is presented in Section VI. Finally, the paper is concluded in Section VII.

E. Bostanci is with Computer Engineering Department, Ankara University, 50. Yil Campus, I Blok, Golbasi, Ankara, 06830, Turkey, e-mail:ebostanci@ankara.edu.tr

II. FINDING POSITION ESTIMATES

Before describing the details of the fusion filter and the DFA approach, it is important to present the calculations used for obtaining individual measurements from the GPS (Phidgets 1040) and IMU sensors (Phidgets Spatial 1056), both low-cost sensors with reasonable accuracy.

A. GPS position estimate

The data obtained from the GPS is in well-known NMEA format and includes position, the number of visible satellites and detailed satellite information for a position P on Earth's surface, as shown in Figure 1.
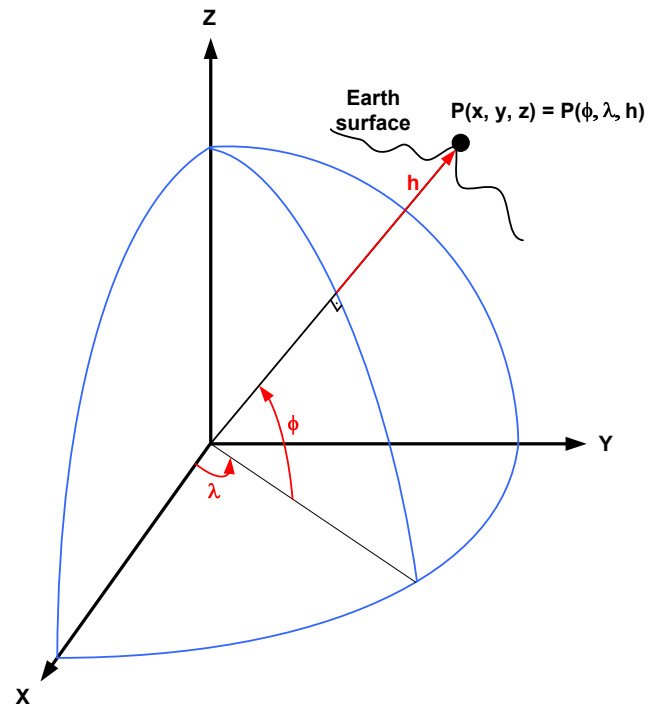

Fig. 1. GPS position parameters in latitude (ϕ), longitude (λ) and altitude (h) and x, y and z in ECEF. Following [17].

Using this information, the GPS coordinates can be converted from geodetic latitude (ϕ), longitude (λ) and altitude (h) notation to ECEF Cartesian coordinates $\mathbf{x_{gps}}$, $\mathbf{y_{gps}}$ and $\mathbf{z_{gps}}$ as:

$$\begin{aligned}
x_{gps} &= (N+h)\cos(\phi)\cos(\lambda) \\
y_{gps} &= (N+h)\cos(\phi)\sin(\lambda) \\
z_{gps} &= ((1-e^2)N+h)\sin(\phi)
\end{aligned} \tag{1}$$

where

$$N = \frac{a}{\sqrt{1.0 - e^2 \sin(\phi)^2}} \qquad (2)$$

and a is the WGS84 [18] ellipsoid constant for equatorial earth radius (6,378,137m), e^2 corresponds to the eccentricity of the earth with a value of $6.69437999 \times 10^{-3}$ [5]. The calculated values form the measurements from the GPS sensor as $m_{gps} = (x_{gps}, y_{gps}, z_{gps})$.

B. IMU position estimate

Finding the position estimate from the IMU is performed by double-integrating the accelerometer outputs for several samples, the current implementation uses four samples. The first integration, to find the velocity, involves integrating accelerations using $v(t) = v(0) + at$:

$$
\begin{aligned}
v_x &= \int_0^T a_x \mathrm{d}t = v_x(T) - v_x(0) \\
v_y &= \int_0^T a_y \mathrm{d}t = v_y(T) - v_y(0) \qquad (3) \\
v_z &= \int_0^T a_z \mathrm{d}t = v_z(T) - v_z(0)
\end{aligned}
$$

Since multiple samples are taken, $\mathrm{d}t$ is the time passed for each one of them. The next step is to integrate the velocities from (3) to find the position using $x(t) = x(0) + vt$ as

$$
\begin{aligned}
x_{imu} &= \int_0^T v_x \mathrm{d}t = p_x(T) - p_x(0) \\
y_{imu} &= \int_0^T v_y \mathrm{d}t = p_y(T) - p_y(0) \qquad (4) \\
z_{imu} &= \int_0^T v_z \mathrm{d}t = p_z(T) - p_z(0)
\end{aligned}
$$

These calculated positions ($m_{imu} = (x_{imu}, y_{imu}, z_{imu})$) are used as the measurements from the IMU.

III. FUSION FILTER

The filter designed for integration of the two sensors consists of a state x which includes positional data ($P = (P_x, P_y, P_z)^T$), linear velocities ($V = (V_x, V_y, V_z)^T$):

$$x = (P, V)^T \qquad (5)$$

A simple state consisting of 6 elements will facilitate obtaining a better performance in speed than one with a larger state. At each iteration, the predict–measure–update cycle of the KF is executed in order to produce a single output from several sensors as the filter output.

In the first stage, *i.e.* prediction, a transition matrix (F of (6)) is applied to the state x in order to obtain the predicted position:

$$
F = \begin{bmatrix}
1 & 0 & 0 & \Delta t & 0 & 0 \\
0 & 1 & 0 & 0 & \Delta t & 0 \\
0 & 0 & 1 & 0 & 0 & \Delta t \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \qquad (6)
$$

where Δt is the time between two prediction stages.

Measurements are obtained from the GPS and the IMU using the values obtained as described in Section II and are combined to create a measurement vector:

$$z = (x_{gps} + x_{imu}, y_{gps} + y_{imu}, z_{gps} + z_{imu})^T \qquad (7)$$

Here, the IMU measurements for position are used as offsets to the position obtained from the most recent GPS fix.

IV. DFA BASED MODEL TRANSITIONS

The difference between the measurements (z) and the prediction ($h\hat{x}$), omitting the subscripts indicating time, is defined as the innovation (y):

$$y = z - h\hat{x} \qquad (8)$$

The innovation vector has 3 components for position elements as y_x, y_y and y_z. The DFA model presented here uses the magnitude of these to define the filter divergence as

$$I = \sqrt{y_x^2 + y_y^2 + y_z^2} \qquad (9)$$

and uses the following rules to assign the values of I into different classes named I_0, I_1 and I_2 which are defined as

$$
I = \begin{cases}
I_0 : & I < 3.0 \\
I_1 : & 3.0 \leq I < 7.5 \\
I_2 : & 7.5 \leq I
\end{cases} \qquad (10)
$$

A DFA consists of several elements which can be listed as states, input symbols and transition rules [19], [20]. The states of the DFA defined in Figure 2 correspond to different motion models.
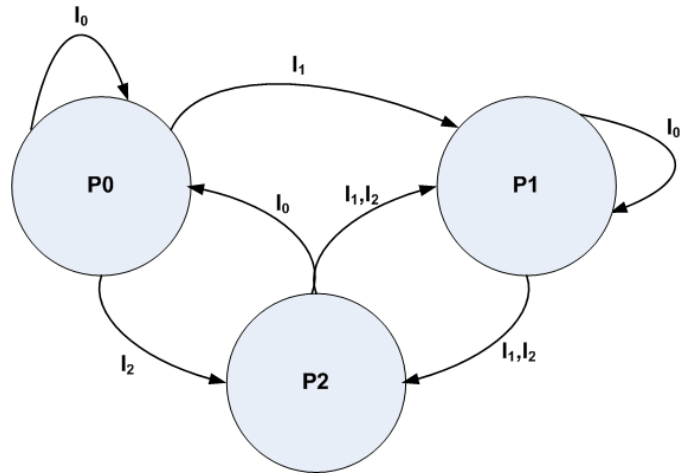

Fig. 2.   DFA model for the model transitions

The classes (I_0, I_1 and I_2) are considered as the input symbols used for the DFA model. Finally, the transitions between states model the selection mechanism presented in this paper.

When a model (P_i) is selected, the value for i is used as a velocity coefficient (c_i) in the transition function (F) for position ($\hat{x}_P = x_P + c_i V \Delta t$). For instance, P0 indicates

a stationary transition model where the current values for position (P) will be unchanged in the predicted state, whereas P2 indicates a motion model where position is predicted with twice the current positional velocities ($\hat{x}_P = x_P + (2 \times V)\Delta t$) in order to adapt any sudden changes in the estimated position.

During experiments it was observed that in some cases selected models could be changing very often. A sliding window filter was applied to the results of the model selection logic in order to prevent frequent transitions between different motion models. In the implementation, the most recent five models were averaged to obtain the final motion model as illustrated in Figure 3.
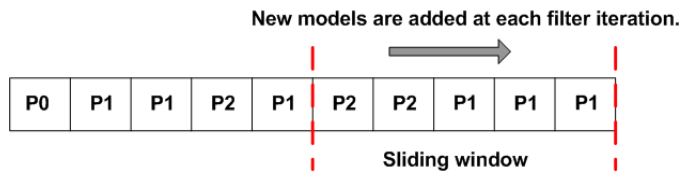
Fig. 3.   Sliding window for preventing frequent model transitions

V. RESULTS

Experiments were conducted using low-cost GPS and IMU sensors mounted on a cycle helmet for a user walking with varying speed. Sampling rate for the IMU was selected as 20 milliseconds and a GPS fix was received every second.

Figures 4 to 8 show the estimated paths using integration of the two sensors and employing different motion models. Portions of the estimated paths are coloured differently in order to indicate the type of the motion model used for estimation. It is important to note that the static model used in the results correspond to P1 and hence is drawn in the same colour.

Dataset 3 was acquired while the sensors were completely stationary, the accuracy of the sensor fusion is found in this case as 1.5m which is, indeed, less than the accuracy of the GPS used in the experiments (given as 2.5m in the product specification) — a benefit of sensor fusion. The DFA model selection logic reduced this error even further since the motion model is correctly recognized as P0 (see Figure 6).

Filter errors are presented in Figures 9 to 13. Note that these errors are an indicator of the difference between the filter predictions and the actual values of the measurements. It can be seen that the filter error is reduced when the DFA models are employed.
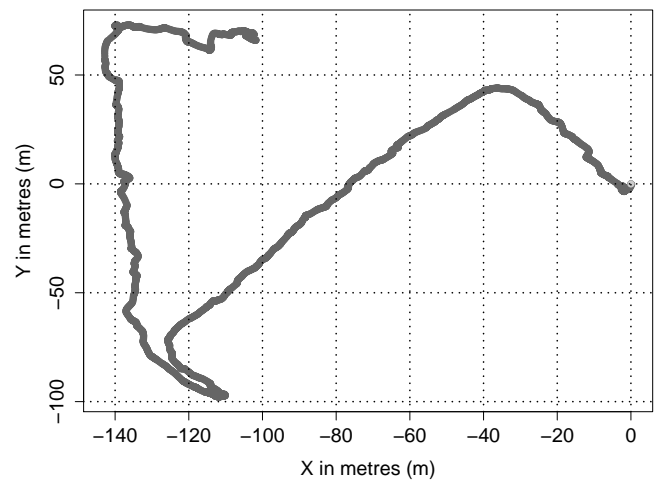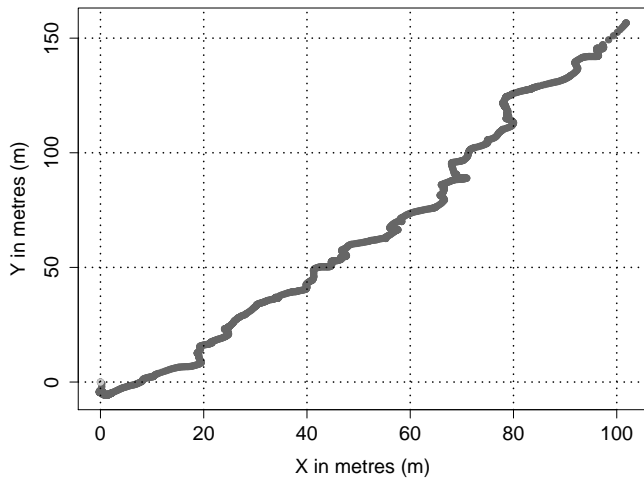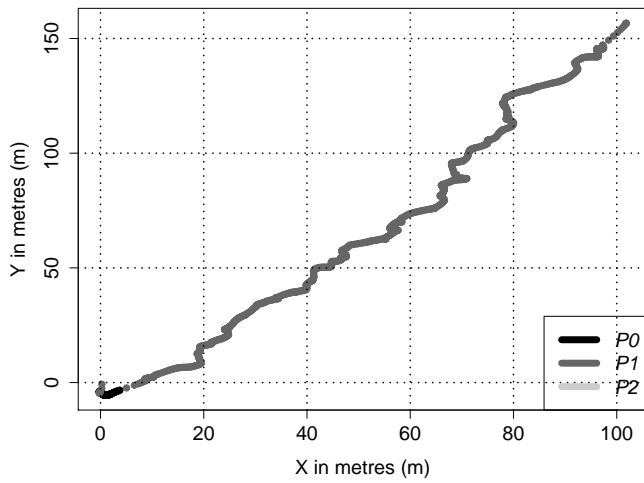

(a)
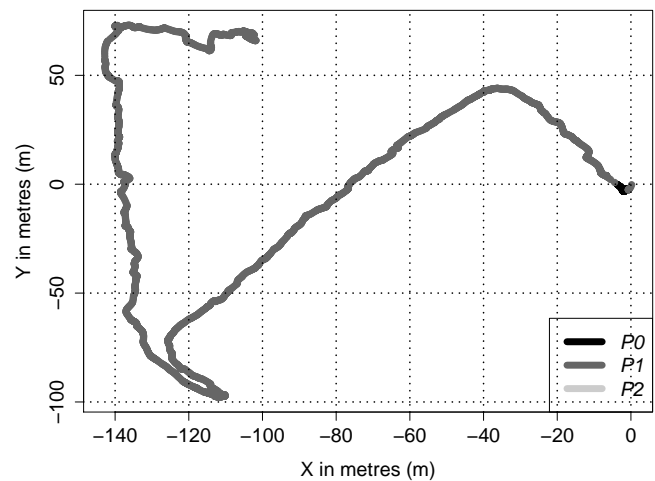
(b)

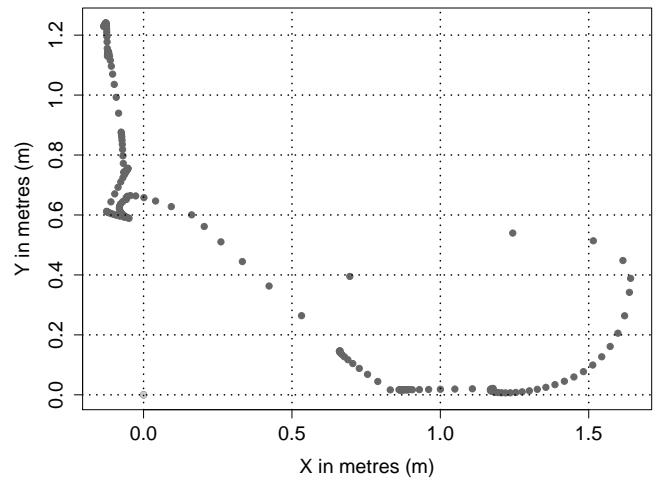Fig. 4.   Trajectory results for Dataset 1. (a) Static motion model (b) DFA models
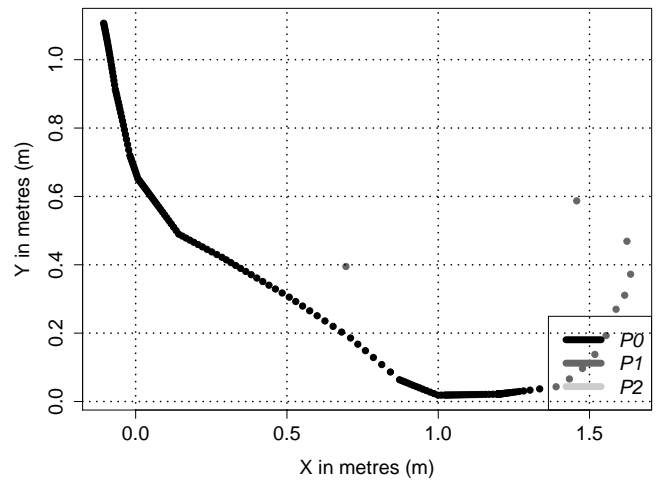
(a)

(b)

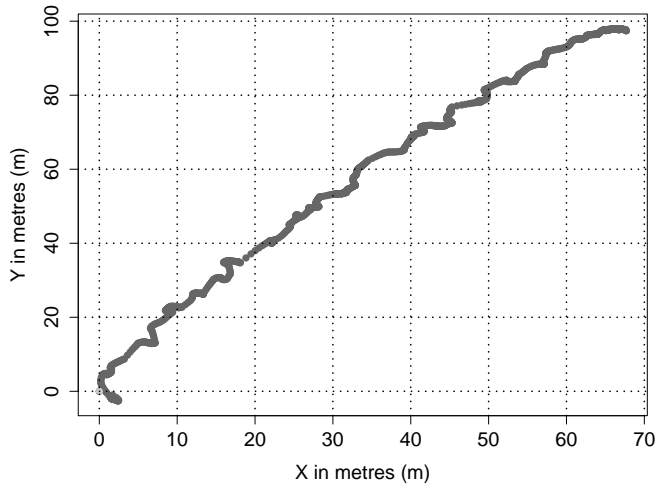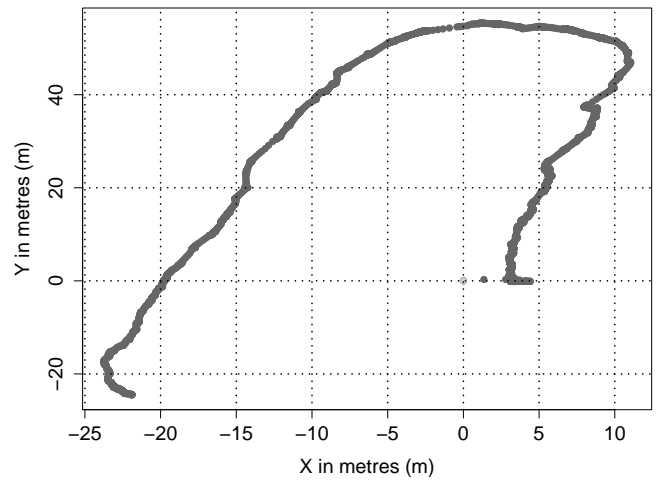Fig. 5.   Trajectory results for Dataset 2. (a) Static motion model (b) DFA models
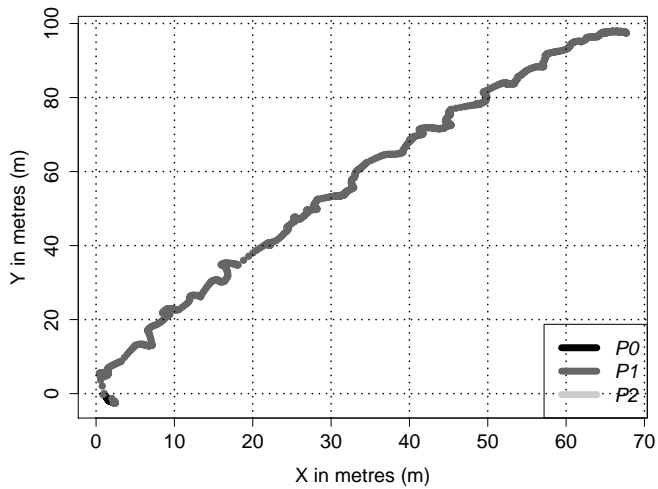

(a)

(b)

Fig. 6.   Trajectory results for Dataset 3. (a) Static motion model (b) DFA models
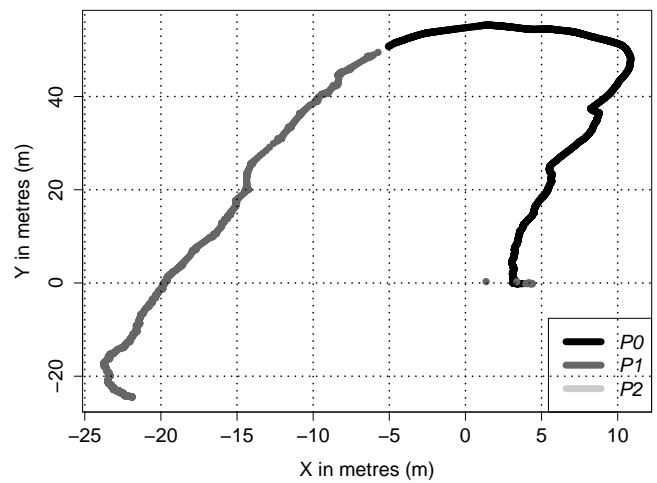
(a)

(b)

Fig. 7.   Trajectory results for Dataset 4. (a) Static motion model (b) DFA models

(a)

(b)

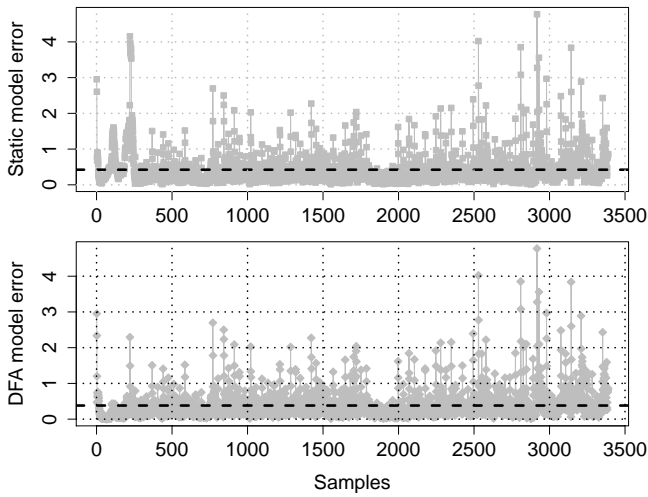Fig. 8.   Trajectory results for Dataset 5. (a) Static motion model (b) DFA models

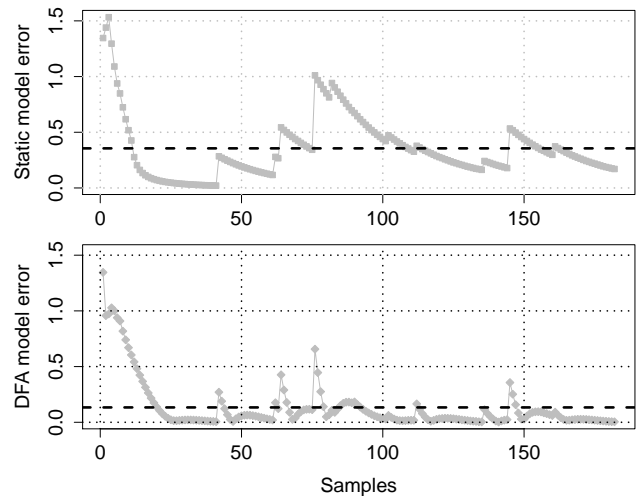Fig. 9. Filter errors for Dataset 1

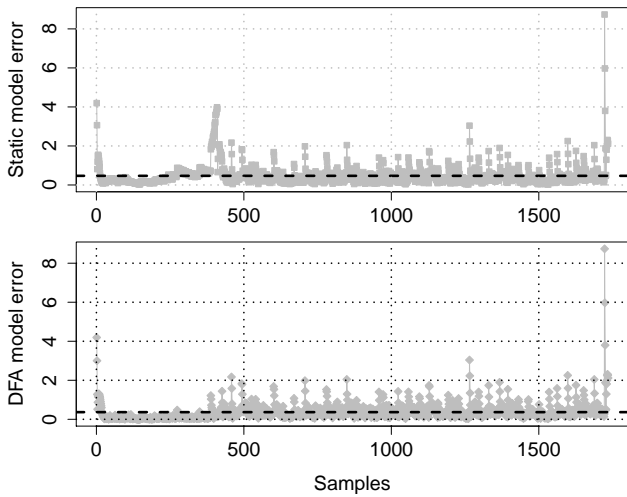Fig. 11. Filter errors for Dataset 3
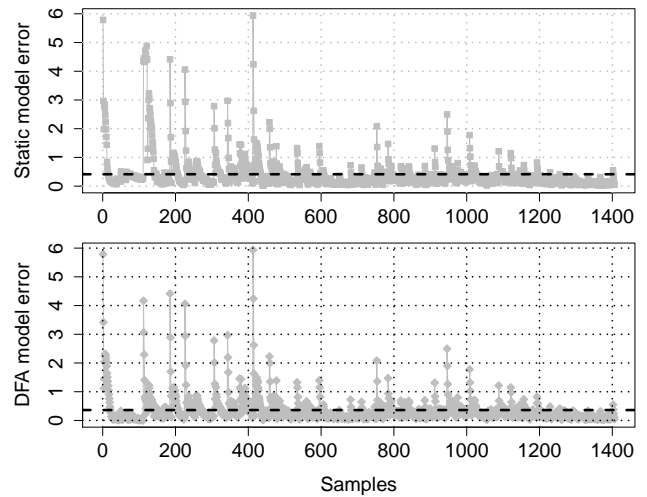

Fig. 10. Filter errors for Dataset 2

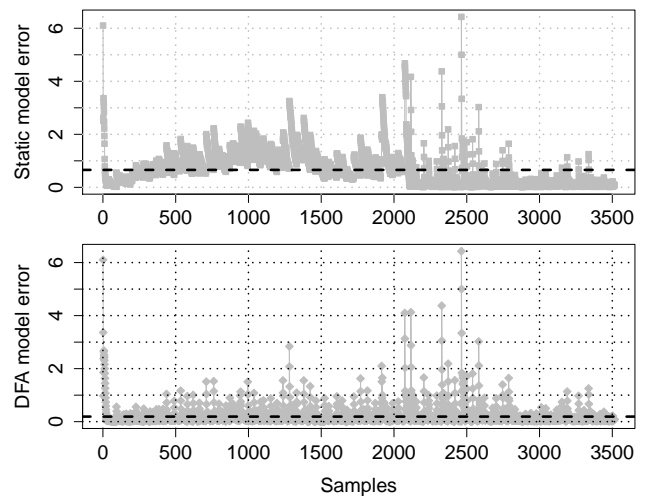Fig. 12. Filter errors for Dataset 4

Fig. 13. Filter errors for Dataset 5

VI. An AR Game – Treasure Hunt

This section presents an AR game, which works using the DFA based sensor fusion algorithm described earlier in the paper. The aim of the game is to collect items and direct the user to test the accuracy of the approach, albeit unconsciously.

The game presents an egocentric view of the environment, as in First Person Shooter (FPS) games. The rules of the game are quite simple: the user needs to reach and collect all the reward items available as quickly as possible. When he or she reaches an item, the score is incremented by an amount that depends on the type of item encountered. The game provides three types of items: small coins, large coins and a chest (Figure 14), with rewards of 10, 30 and 50 points respectively.

(a) Chest   (b) Coin

Fig. 14.   Models used in the AR game

After the game is initialized with the positions of all items set, the game loop starts. The coin models use the animator and they rotate about their axes while the chest models remain static.

At each frame, the position of the user is checked against the item positions by calculating the distance between them. If this distance is less than some threshold value (done so that there is some tolerance against positioning inaccuracies), then the score is updated, the item is set as 'hit' and a sound file is played. The items collected by the user simply disappear. A timer is used for two purposes. First, it is constantly updated in the display to provide feedback to the user. It is also used to decay the score

$$score = initialScore \times c/time \qquad (11)$$

where $score$ is the final score to be added and $initialScore$ correspond to the rewards mentioned above. The constant c is selected as 5.0 arbitrarily. This forces the user to collect the game tokens quickly.

A view from the AR game is presented in Figure 15. The game has an interface which displays the score and time passed making the game more challenging and hence interesting. Note that the frame rate of the game is, indeed, very close to video rates (22 frames per second), an indicator of the speed of the filter.

Fig. 15.   A view from the AR game

VII. Conclusion

This paper presented a DFA design for motion model selection in GPS–IMU sensor fusion. The results show that multiple-motion model sensor fusion can be achieved by utilising Kalman filter innovation together with a DFA based model selection scheme. It was observed that the use of different motion models can reduce the filter error and prevent divergence. It is clear that choosing the appropriate motion model depending on user's speed improves the accuracy of Kalman filter for tracking applications.

A sample AR game was used to test the defined approach, and it was observed that the filter is accurate and fast enough to collect all the reward items in the game.

Future work will delve into further analysis of different motion models and a machine learning approach appears to be a promising research direction.

References

[1] P. Groves, *Principles of GNSS, inertial, and multi-sensor integrated navigation systems*, ser. GNSS technology and applications series. Artech House, 2008.

[2] T. L. Grigore, R. M. Botez, D. G. Sandu, and O. Grigorie, "Experimental testing of a data fusion algorithm for miniaturized inertial sensors in redundant configurations," in *Proceedings of the 2014 International Conference on Mathematical Methods, Mathematical Models and Simulation in Science and Engineering*, 2014, pp. 116–122.

[3] V. Barrile and G. Bilotta, "Automated surveys and integrated auto-location by laser scanner and gps," in *Proceedings of the 2014 International Conference on Communications, Signal Processing and Computers*, 2013, pp. 65–72.

[4] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME Journal of Basic Engineering*, no. 82 (Series D), pp. 35–45, 1960.

[5] E. Kaplan and C. Hegarty, *Understanding GPS: Principles and Applications*, ser. Artech House mobile communications series. Artech House, 2005.

[6] A. Almagbile, J. Wang, and W. Ding, "Evaluating the performances of adaptive kalman filter methods in GPS/INS integration," *Journal of Global Positioning Systems*, vol. 9, no. 1, pp. 33–40, 2010.

[7] C. Tseng, C. Chang, and D. Jwo, "Fuzzy adaptive interacting multiple model nonlinear filter for integrated navigation sensor fusion," *Sensors*, vol. 11, pp. 2090–2111, 2011.

[8] J. Kramer and A. Kandel, "On accurate localization and uncertain sensors," *International Journal of Intelligent Systems*, vol. 27, no. 5, pp. 429–456, 2012.

[9] L. Ojeda and J. Borenstein, "Flexnav: fuzzy logic expert rule-based position estimation for mobile robots on rugged terrain," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 1, 2002, pp. 317–322 vol.1.

[10] S. K. Hong, "Fuzzy logic based closed-loop strapdown attitude system for unmanned aerial vehicle (uav)," *Sensors and Actuators A: Physical*, vol. 107, no. 2, pp. 109 – 118, 2003.

[11] J. Chen and A. Pinz, "Structure and motion by fusion of inertial and vision-based tracking," in *Austrian Association for Pattern Recognition*, vol. 179, 2004, pp. 75–62.

[12] P. Torr, "Bayesian model estimation and selection for epipolar geometry and generic manifold fitting," *International Journal of Computer Vision*, vol. 50, no. 1, pp. 35–61, 2002.

[13] K. Kanatani, "Uncertainty modeling and model selection for geometric inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 10, pp. 1307–1319, 2004.

[14] K. Schindler and D. Suter, "Two-view multibody structure-and-motion with outliers through model selection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 6, pp. 983–995, 2006.

[15] J. Civera, A. Davison, and J. M. M. Montiel, "Interacting multiple model monocular slam," in *International Conference on Robotics and Automation*, 2008, pp. 3704–3709.

[16] E. Bostanci, "A DFA approach for motion model selection in sensor fusion," in *Proceedings of the 2014 International Conference on Neural Networks - Fuzzy Systems*, 2014, pp. 53–57.

[17] S. H. Stovall, "Basic inertial navigation," Naval Air Warfare Center Weapons Division, Technical Report, 2008.

[18] National Imagery and Mapping Agency, "Department of Defense World Geodetic System 1984: its definition and relationships with local geodetic systems," National Imagery and Mapping Agency, Tech. Rep., 2000.

[19] J. Hopcroft, R. Motwani, and J. Ullman, *Introduction to automata theory, languages, and computation*. Pearson/Addison Wesley, 2007. [Online]. Available: http://books.google.com.tr/books?id=6ytHAQAAIAAJ

[20] Y. Uchida, T. Ito, M. Sakamoto, R. Katamune, K. Uchida, H. Furutani, M. Kono, S. Ikeda, and T. Yoshinaga, "Path-bounded finite automata on four-dimensional input tapes," *International Journal of Computers*, vol. 1, no. 5, pp. 58–65, 2011.

Erkan Bostanci received a BSc degree in Computer Engineering department from Ankara University, Turkey in 2007. Consequently, he joined the same department as a Research Assistant and completed his MSc on real-time battlefield simulation in 2009. He obtained his PhD from School of Computer Science and Electronic Engineering, University of Essex, UK in 2014.

His research interests include different yet closely related aspects of computer science from image processing, computer vision and graphics to artificial intelligence and fuzzy logic as well as mathematical modelling and statistical analysis. He developed a vision-based user tracking system for augmented reality applications for cultural heritage in particular.

Dr. Bostanci is a member of the IEEE and he has been involved in technical committees for several conferences and acted as a reviewer for various journals.