

# Kalman Filter for Hybrid Tracking Technique in Augmented Reality

Edgaras Artemciukas<sup>1,2</sup>, Leonidas Sakalauskas<sup>1,3</sup>, Evaldas Zulkas<sup>4</sup>

<sup>1</sup>*Department of System Analysis, Institute of Mathematics and Informatics, Vilnius University, Akademijos St. 4–319, LT-08412 Vilnius, Lithuania*

<sup>2</sup>*Department of Electrical and Mechanical Engineering, Faculty of Technology, Klaipeda State College, Bijunu St. 10–312, LT-91223 Klaipeda, Lithuania*

<sup>3</sup>*Department of Information Technology, Faculty of Fundamental Sciences, Vilnius Gediminas Technical University,*

*Sauletekio avenue 11, LT-10223 Vilnius, Lithuania*

<sup>4</sup>*Department of Software Engineering, Institute of Mathematics and Informatics, Vilnius University, Akademijos St. 4, LT-08412 Vilnius, Lithuania*  
*edgaras.artemciukas@mii.vu.lt*

**Abstract**—Augmented reality started to emerge as a promising visualization technique that tracks real objects and adds virtual content into real world context using camera view. Many augmented reality solutions are based on computer vision techniques to identify and track objects. Problems that must be solved are image transformations, chaotic environment, lighting condition and occlusion from users' or objects in the environment, which causes virtual content to disappear. This has a negative impact for augmented reality usability, therefore, object recognition and tracking in real-time becomes difficult and sometimes an impossible task. In this research orientation-position information acquisition using computer vision and sensor fusion techniques are analysed. Experiments are accomplished with predefined assumptions and simulated orientation-position information. Conditions for optimal orientation-position estimates are introduced. Research results are compared and supplementary properties are presented of a proposed hybrid tracking technique using Kalman filter.

**Index Terms**—Augmented reality; computer vision; hybrid tracking; Kalman filter; sensor fusion.

## I. INTRODUCTION

Most of the research work in augmented reality (AR) field investigates problems that are related to tracking and interaction techniques. Using tracking techniques virtual content can be represented in appropriate orientation and position, while changing users' perspective, and interaction techniques allow users to manipulate with virtual content. Nowadays implemented object tracking solutions for AR are often based on computer vision (CV) techniques such as SIFT, FERNS, SURF, FAST or other similar methods and their modifications [1]–[5]. Several tasks must be taken into consideration using augmented reality systems that are based on computer vision techniques:

1. Locate and track object in the scene.
2. Display virtual content depending on a trackable object orientation and position.
3. Ensure an opportunity to interact with content.

While tracking object in real environment conditions several problems must be solved: illumination level, image transformations because of different camera perspective, image quality, reflection and partial or full occlusion at the same time. As a result, less permanent features are detected and matches are found between different viewpoints of the same scene to accomplish reliable object tracking. Virtual content is not displayed, if the tracking is lost. Image processing speed is also critical aspect for AR and must be accomplished in real-time.

Depending on AR application field ultrasonic tracking technique [6] is a solution for position tracking. However, it is limited to workspace and no orientation estimates are provided. Currently object motion tracking using digital inertial sensors is an active research topic that are analysed in [7], [8] works. Fast and irregular camera movement causes tracking errors and instabilities in case of computer vision. These problems can be solved using inertial sensors to estimate rapid camera orientation changes. Camera tracking using inertial sensors is a suitable method because of a high speed measurement acquisition. However, it is important to maintain stability after a longer period of time. Sensors are affected by noise, drift and magnetic interference. By integrating and combining several sensor information disadvantages of separate sensors can be eliminated using sensor fusion solutions [9]–[11]. In this way reliable orientation estimates are provided.

This paper describes position and orientation estimation problems, which is critical in the field of augmented reality, and proposes a way to improve it. In this research no specific computer vision method was proposed. However, assumptions were made that object can be recognized and tracked in an image at 20 frames per second (has a disadvantage in speed). Wrong or unavailable orientation-position estimates using computer vision tracking can be improved or supplemented with additional estimates using sensors. Orientation and position estimation is the main aspect, which is analysed in this work. To achieve better

orientation-position estimates Kalman filter for hybrid tracking technique was designed. Computer vision (CV) and sensor fusion (SF) tracking information were simulated.

## II. OBJECT TRACKING USING SENSOR FUSION

Orientation estimation using digital sensors is an active research topic. Digital accelerometer measures acceleration and gravity of the device. In general case, this sensor is suitable to estimate an orientation of the object, depending on Earth's gravitational force. According to accelerometer acquired measurements, orientation  $a_{\phi,\theta} = [a_\phi, a_\theta]^T$  vector can be estimated by using (1)

$$a_{\phi,\theta} = \begin{bmatrix} a_\phi \\ a_\theta \end{bmatrix} = \begin{bmatrix} \arctan\left(a_y / \sqrt{a_x^2 + a_z^2}\right) \\ \arctan\left(a_x / \sqrt{a_y^2 + a_z^2}\right) \end{bmatrix}, \quad (1)$$

where  $a_x, a_y, a_z$  – accelerometer output at current time moment with respect to  $x, y$  and  $z$  axis. Sensor measurements in a stable state are affected by noise, therefore,  $a_\phi$  (rotation around  $x$  axis – roll) and  $a_\theta$  (rotation around  $y$  axis – pitch) orientation estimates can be wrong and unreliable. Gravity force does not provide information about  $a_\psi$  angle (rotation around  $z$  axis – yaw).

Gyroscope angular rate measurements are not affected by noise and do not accumulate errors. Object orientation vector  $\omega_{\phi,\theta,\psi} = [\omega_\phi, \omega_\theta, \omega_\psi]^T$  is estimated by using (2)

$$\omega_{\phi,\theta,\psi} = \begin{bmatrix} \omega_{\phi_t} \\ \omega_{\theta_t} \\ \omega_{\psi_t} \end{bmatrix} = \begin{bmatrix} \omega_{\phi_{t-1}} + \omega_{x_t} \Delta t \\ \omega_{\theta_{t-1}} + \omega_{y_t} \Delta t \\ \omega_{\psi_{t-1}} + \omega_{z_t} \Delta t \end{bmatrix}, \quad (2)$$

where  $\omega_{x_t}, \omega_{y_t}, \omega_{z_t}$  – gyroscope angular rate at current time moment  $t$ ;  $\omega_{\phi_{t-1}}, \omega_{\theta_{t-1}}, \omega_{\psi_{t-1}}$  – estimated orientation at earlier time moment  $t-1$ ;  $\Delta t$  – time between measurements. Estimated orientation using gyroscope angular rate measurements are accurate only in a short period of time. After a longer period of time drift exponentially increases and errors are accumulated without any fixed reference system. For augmented reality tracking system, it is a critical aspect.

Heading  $b_\psi$  from magnetometer can be determined using the following (3)–(5):

$$X_h = m_x \cos \theta + m_y \sin \phi \sin \theta + m_z \cos \phi \sin \theta, \quad (3)$$

$$Y_h = m_y \cos \phi \sin \theta - m_z \sin \phi, \quad (4)$$

$$b_\psi = -\arctan(Y_h/X_h), \quad (5)$$

where  $m_x, m_y, m_z$  – magnetic field measurements from magnetometer;  $\phi, \theta$  corresponds to  $a_\phi$  and  $a_\theta$  estimated orientation using the accelerometer (1). One of the magnetometers disadvantages – magnetic field measurements are affected by distortions, which are caused by ferromagnetic objects. These distortions can be compensated.

A combination of an accelerometer, gyroscope and magnetometer digital sensors measurements are used to ensure reliable object orientation in 3D space. Orientation drift using gyroscope is eliminated with accelerometer, which contributes in correcting  $\phi$  and  $\theta$  angles. Magnetometer ensures corrected  $\psi$  heading measurements in combination with accelerometer. This is the main idea for sensor fusion using quaternion representation. Quaternion requires less calculation time, provides reliable orientation estimation and maintains stability compared to Euler angles or rotation matrix representation. Quaternion  $q$  is a four element vector (6)

$$q = [q_0, q_1, q_2, q_3] = [w, xi, yj, zk], \quad (6)$$

where  $q_0(w)$  determines the rate of rotation;  $q_1(x), q_2(y), q_3(z)$  – rotations in respect to  $x, y$  and  $z$  axis. Quaternion-based algorithms for orientation estimation using sensors are explicitly analysed in [8], [9], therefore, no detailed analysis are provided in this work. Systems that uses such orientation estimation techniques are not limited to motion, specific environment, place or occlusions, therefore, it has advantages over computer vision techniques.

Even though quaternion representation of orientation is more reliable, for simplicity purposes in further hybrid tracking experiments Euler angles are used instead of quaternions. Euler angles  $e_a = [\phi, \theta, \psi]^T$  can be converted from quaternions using (7) expression

$$e_a = \begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} \tan^{-1}\left(\frac{2q_2q_3 - 2q_0q_1}{2q_0^2 + 2q_3^2 - 1}\right) \\ -\sin^{-1}(2q_1q_3 + 2q_0q_2) \\ \tan^{-1}\left(\frac{2q_1q_2 - 2q_0q_3}{2q_0^2 + 2q_1^2 - 1}\right) \end{bmatrix}. \quad (7)$$

Position estimation is a difficult, but possible task using accelerometer measurements data. As mentioned before, accelerometer output jointly measures acceleration and gravity, therefore, it is necessary to eliminate gravity to get linear acceleration. If measurements from sensor array (accelerometer, gyroscope and magnetometer) are available, then gravity vector  $g = [g_x, g_y, g_z]^T$  can be estimated using orientation represented in quaternion (8)

$$g = \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} = \begin{bmatrix} 2q_1q_3 - 2q_0q_2 \\ 2q_0q_1 + 2q_2q_3 \\ q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}, \quad (8)$$

where  $q_0, q_1, q_2, q_3$  – quaternion elements that represent orientation;  $g_x, g_y, g_z$  – accelerometer gravity direction with respect to each axis. Estimated orientation using quaternion representation must be accurate. Even small errors in orientation estimation that are used to calculate gravity vector can cause large errors in linear acceleration. Linear acceleration vector  $a_g$  with eliminated gravity  $g$  from

accelerometer output can be estimated using (9)

$$a_g = \begin{bmatrix} a_{g_x} \\ a_{g_y} \\ a_{g_z} \end{bmatrix} = \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} - \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix}. \quad (9)$$

Using linear acceleration  $a_g$ , velocity  $v_t$  and position  $p_t$  vectors are estimated for each axis using (10) and (11) expressions:

$$v_t = v_{t-1} + a_{g_t} \Delta t, \quad (10)$$

$$p_t = p_{t-1} + v_t \Delta t + a_{g_t} \Delta t^2 / 2, \quad (11)$$

where  $\Delta t$  – time between measurements. As accelerometer measurements are affected by noise, it cannot be efficiently eliminated for position estimation. For further experiments estimated orientation and position-velocity-acceleration vector  $x_{sf}$  using sensors are modelled using (12) expression

$$x_{sf} = \left[ \begin{array}{l} \varphi_{sf}, \theta_{sf}, \psi_{sf}, [p_x, v_x, a_{g_x}], \\ [p_y, v_y, a_{g_y}], [p_z, v_z, a_{g_z}] \end{array} \right]. \quad (12)$$

### III. OBJECT TRACKING USING COMPUTER VISION

Computer Vision methods use feature extraction (detection) and description as a primary analysis aspect to find interest points in an image and matching to perform object tracking. This allows estimating relative orientation and position of the camera using RANSAC probability method. From geometric perspective it can be explained by using pinhole camera model, which is widely used and analysed in computer vision researches. It defines a relationship between 3D point  $P = [X, Y, Z]^T$  from the scene and 2D corresponding projection  $p = [x, y]^T$  onto the image plane. Mapping from 3D to 2D is called perspective projection and can be expressed using (13):

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}}_{\text{intrinsic parameter}} \underbrace{\begin{bmatrix} R_{11} & R_{12} & R_{13} & T_1 \\ R_{21} & R_{22} & R_{23} & T_2 \\ R_{31} & R_{32} & R_{33} & T_3 \end{bmatrix}}_{\text{extrinsic parameter}} \underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_P, \quad (13)$$

where  $f_x, f_y$  – focal length;  $c_x, c_y$  – optical centre of the camera. Intrinsic parameters are used to remove distortions. The extrinsic parameter is a transformation matrix in camera coordinate system, which consists of rotation matrix  $R_{3 \times 3}$  for orientation and translation vector  $T_{3 \times 1}$  for position. Computer vision techniques focus on estimating this fundamental transformation matrix. Rotation matrix can be converted to Euler angles using (14) expression:

$$e_a = \begin{bmatrix} \varphi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} \tan^{-1}(R_{23}/R_{33}) \\ -\sin^{-1}(R_{13}) \\ \tan^{-1}(R_{12}/R_{11}) \end{bmatrix}. \quad (14)$$

In case of computer vision, orientation and position vector  $x_{cv}$  is modelled using (15)

$$x_{cv} = [\phi_{cv}, \theta_{cv}, \psi_{cv}, p_x, p_y, p_z]. \quad (15)$$

where  $\varphi_{cv}, \psi_{cv}, \theta_{cv}$  converted orientation from rotation matrix  $R_{3 \times 3}$  to Euler angles using (14) and  $p_x, p_y, p_z$  are position coordinates from translation vector  $T_{3 \times 1}$  in (13).

### IV. KALMAN FILTER FOR HYBRID TRACKING

Two main aspects are important from user's perspective using AR system:

1. System must operate in real-time without any delay.
2. Even in rapid movement and occlusion tracking must be robust without any jittering.

While representing virtual content in augmented reality environment these aspects must be ensured with no interruption using hybrid tracking. Dynamic motion measurements from sensors are used to improve provided information from computer vision method. Conceptual diagram of data acquisition and processing from sensors and general purpose camera is presented in Fig. 1.

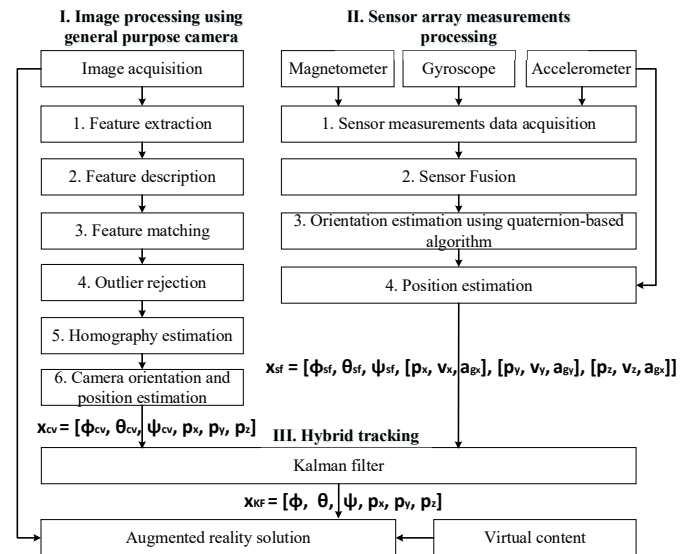


Fig. 1. Kalman filter for Hybrid tracking: orientation and position estimation using computer vision and sensor fusion techniques.

This is a general model for hybrid tracking technique. The following assumptions for further hybrid tracking technique must be taken into consideration:

1. Calibration of camera and sensors are accomplished offline.
2. Sensors are rigidly attached to camera, therefore, it is considered that coordinate systems coincide.
3. Data are acquired and processed synchronized to get orientation-position information by applying CV and SF

methods. Even small synchronization misalignments can lead to large tracking errors.

4. CV and SF data processing are implemented on different threads.

5. Processed data acquisition frequency from CV method is  $f_{cv} = 20$  Hz ( $\Delta t = 0.05$ ) and SF –  $f_{sf} = 100$  Hz ( $\Delta t = 0.01$ ).

6. Sensors attached to the camera moves in the environment, not the object that can be tracked in the scene using CV.

7. CV method provides relative information about camera orientation-position. Accordingly, CV and SF orientation-position information are simulated from the common starting point.

The last assumption is the main aspect for orientation-position estimation, because information does not coincide when  $x_{sf}$  and  $x_{cv}$  vectors are taken separately. For instance, in case of CV relative camera orientation-position information is estimated from the object that is found and tracked in the scene and sensor fusion solution provides this information directly from a camera motion.

After taking assumptions into consideration, Kalman filter (KF) is applied for simulated orientation-position information. KF algorithm is recursive and widely used in object trajectory prediction, control, tracking, collision-warning systems, image processing, sensor fusion etc. KF consists of prediction (process model) (16)–(17) and update (measurement model) (18–20) steps:

1. State prediction  $\hat{x}_k$  (16)

$$\hat{x}_k = A\bar{x}_{k-1} + w_k. \quad (16)$$

2. State covariance  $\hat{P}_k$  prediction (17)

$$\hat{P}_k = A\bar{P}_{k-1}A^T + Q. \quad (17)$$

3. Gain  $K_k$  calculation to correct state prediction  $\hat{x}_k$  (18)

$$K_k = \hat{P}_k H^T (H\hat{P}_k H^T + R)^{-1}. \quad (18)$$

4. State estimate  $\bar{x}_k$  update using measurement  $w_k$  (19)

$$\bar{x}_k = \hat{x}_k + K_k (z_k - H\hat{x}_k). \quad (19)$$

5. Covariance  $\bar{P}_k$  update (20)

$$\bar{P}_k = (I - K_k H) \hat{P}_k, \quad (20)$$

where  $\hat{x}_k$  – state prediction vector affected by noise  $w_k$ ;  $\bar{x}_k$  – update vector;  $z_k = H\hat{x}_k + v_k$  – measurement vector affected by noise  $v_k$ ;  $w_k \sim N(0, Q)$   $v_k \sim N(0, R)$  – respectively process prediction and measurement update independent Gaussian noises;  $A$  – state transition matrix;  $\hat{P}_k$ ,  $\bar{P}_k$  – respectively prediction and update state covariance matrix;  $Q = E[w_k w_k^T]$ ,

$R = E[v_k v_k^T]$  respectively independent process and measurement noise covariance matrices;  $H$  – measurement matrix;  $I$  – identity matrix. In the update step the difference between measurement and prediction states are compensated and new estimates determined. KF convergence rate depends on  $Q$  and  $R$  values; decreased value of  $Q$  or  $R$  shows confidence in either process or measurement steps.

Analysing independent innovation  $\hat{y}_k$  (21) or residual  $\bar{y}_k$  (22) sequences KF efficiency can be determined, which is a reliable quality indicator:

$$\hat{y}_k = z_k - H_k \hat{x}_k, \quad (21)$$

$$\bar{y}_k = z_k - H_k \bar{x}_k. \quad (22)$$

In an ideal case, dynamic systems' innovation  $\hat{y}_k$  and residual  $\bar{y}_k$  has to be zero or near zero mean value.

## V. EXPERIMENTAL RESULTS

For simplicity purposes SF vector  $x_{sf}$  is separated into two parts: orientation  $x_{sf}^o = [\phi_{sf}, \theta_{sf}, \psi_{sf}]$  and position-velocity-acceleration  $x_{sf}^p = \begin{bmatrix} p_x, v_x, a_{g_x} \\ p_y, v_y, a_{g_y} \\ p_z, v_z, a_{g_z} \end{bmatrix}$  vectors. CV vector  $x_{cv}$  is also separated in orientation  $x_{cv}^o = [\phi_{cv}, \theta_{cv}, \psi_{cv}]$  and position  $x_{cv}^p = [p_x, p_y, p_z]$  parts. From here some matrices are denoted by superscript “o” for orientation and superscript “p” for position or position-velocity-acceleration vectors to avoid confusion:

$$A^o = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (23)$$

$$A^p = \begin{bmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}, \quad (24)$$

$$H^o = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (25)$$

$$H^p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (26)$$

Two different KF perspectives are used for orientation and position estimation. CV has slower data processing rate, therefore, available faster SF information is used for prediction between CV samples. According to the described orientation and position estimation scenarios KF is applied using prediction and update equations (16-20). In orientation estimation scenario three orientation states  $\phi$ ,  $\theta$ ,  $\psi$  are estimated using KF for hybrid tracking. Simulated camera orientation ( $\phi$  angle) using SF and CV are presented in Fig. 2.

Initial tracked object orientation is 90 degrees. Processed

data are acquired at constant time intervals from camera and sensors. Similar orientation results would be obtained for  $\theta$  and  $\psi$  angles. Camera orientation is simulated in rapid motion. Even though most of the time orientation using CV is available, in some cases, tracking is improperly estimated that causes spikes or it can be lost (red crosses) because of the occlusion.

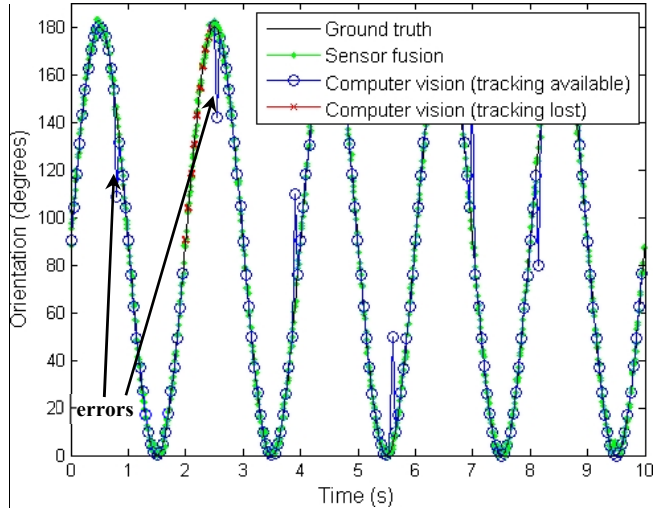


Fig. 2. Simulated SF and CV orientation with modelled outliers and unavailable orientation segment using CV.

Kalman filter for hybrid tracking starts after the orientation information is acquired using CV method and used as an initial state  $\bar{x}_0$  (starting point) in prediction equation. Update vector  $\bar{x}_{k-1}$  in prediction (16) corresponds to  $x_{cv}^o$  computer vision orientation vector and measurement  $z_k$  corresponds to  $x_{sf}^o$  sensor fusion orientation vector. In general case, better confidence is assigned to available CV information ( $Q$  for process), rather than SF ( $R$  for measurement):

$$Q = \begin{bmatrix} 10^{-4} & 0 & 0 \\ 0 & 10^{-4} & 0 \\ 0 & 0 & 10^{-4} \end{bmatrix}, \quad (27)$$

$$R = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{bmatrix}. \quad (28)$$

Posterior update state  $\bar{x}_k$  using SF measurement  $z_k$  is used in prediction step with the same confidence as CV information, if orientation vector  $\bar{x}_{k-1}$  using CV is unavailable for state prediction  $\hat{x}_k$ . Also, the following conditions are introduced to determine the confidence, when wrong CV orientation vector is acquired:

$$\begin{cases} |\hat{y}_k| < z_k \times z_{TH}, & \text{trust is given to process,} \\ |\hat{y}_k| > z_k \times z_{TH}, & \text{trust is given to measurement,} \end{cases} \quad (29)$$

where  $z_{TH}$  – determined 5 % threshold parameter. Innovation comparison with  $z_k$  takes place as measurements from SF are always available. If the error in CV is critical, innovation

exceeds a threshold of  $z_k$  measurement, therefore, better confidence is provided to measurement. In real-life conditions, existing errors in CV provides considerable differences comparing to SF orientation information. If CV information is available without any errors, innovation does not exceed a threshold, as it provides similar results to SF orientation. Estimated orientation using Kalman filter for hybrid tracking presented in Fig. 3.

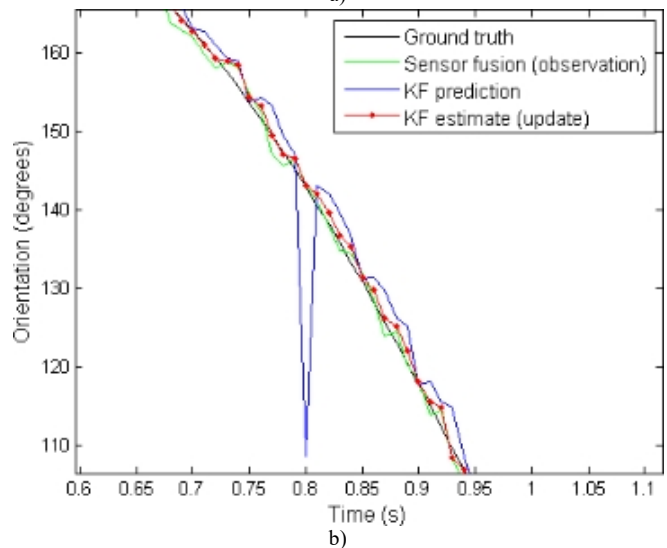
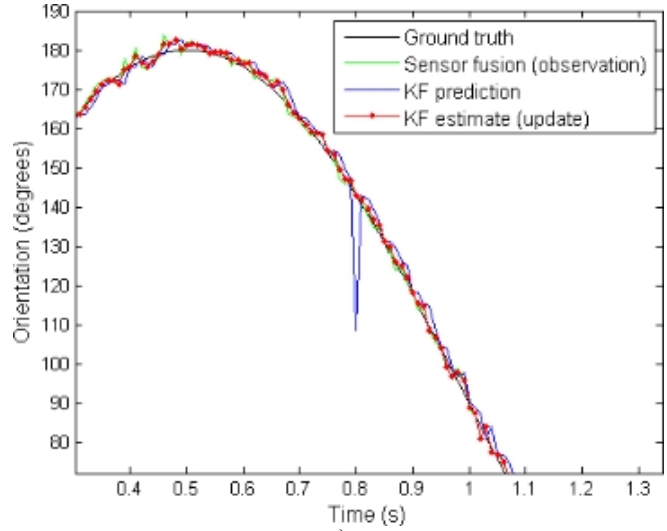


Fig. 3. Hybrid tracking for orientation estimation.

Spike in CV orientation is successfully eliminated using KF. The same applies to unavailable orientation information from CV. Innovation results in case of orientation are presented in Fig. 4.

Spikes in innovation are errors from CV orientation information and are taken into consideration while estimating orientation using Kalman filter for hybrid tracking.

In position estimation scenario only one coordinate is tracked. For position estimation SF information is used in process model and CV information in measurement model.  $A^P$  (23) also known as constant acceleration model (for process model) is used to calculate only 1-axis vector of position, velocity and acceleration. For the other two axes the same model should be applied as well. In case of measurement part, interest holds only in position. Velocity and acceleration are not observed, therefore  $H^P$  is used

according to (26) equation. Simulated noisy acceleration and estimated position are presented in Fig. 5.

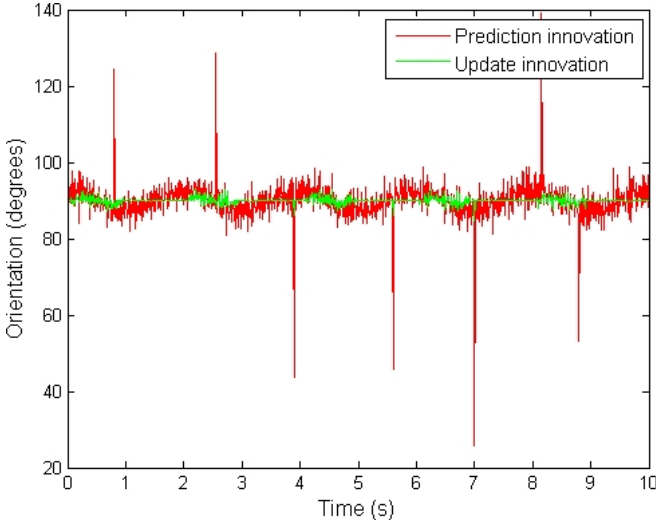


Fig. 4. Orientation prediction and update innovation sequences.

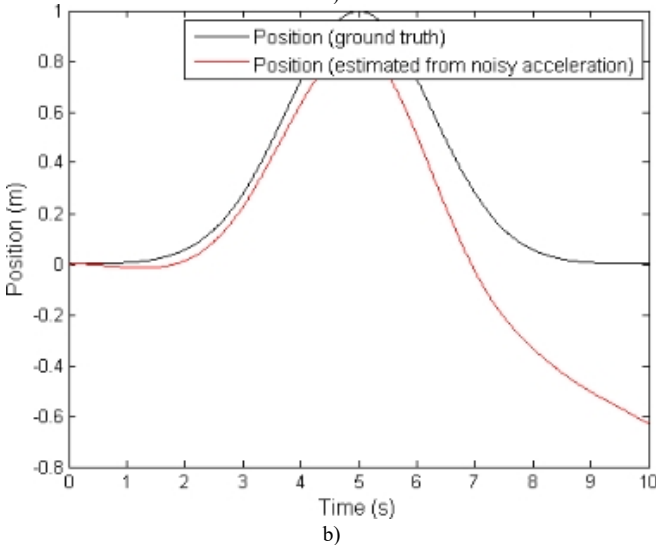
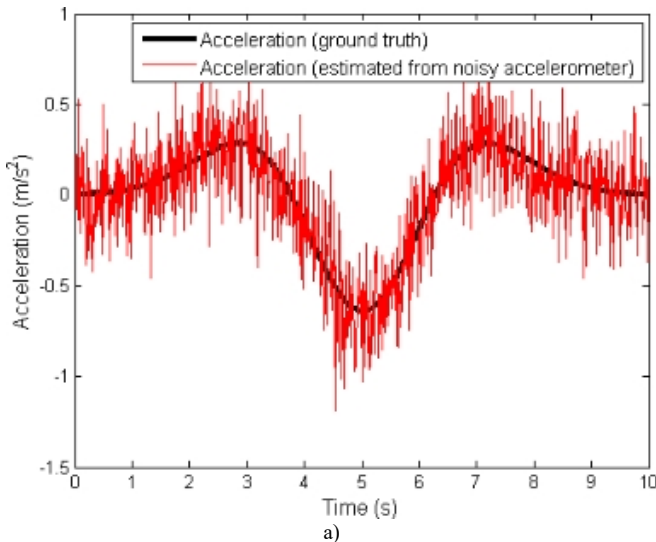


Fig. 5. Simulated acceleration for position estimation: a) noisy acceleration (normal distribution with  $\sigma = 0.2$ ) for position estimation; b) drifted position estimates from noisy acceleration compared to ground truth.

Positive and negative values in the diagrams show a direction of acceleration and position. In process model ( $\bar{x}_{k-1}$ ) position-velocity-acceleration information is

estimated from acceleration and in measurement model ( $z_k$ ) position is acquired from CV. Better confidence is provided to a computer vision information since using noisy acceleration measurements position estimates accumulates considerable errors. Position innovation sequences are presented in Fig. 6.

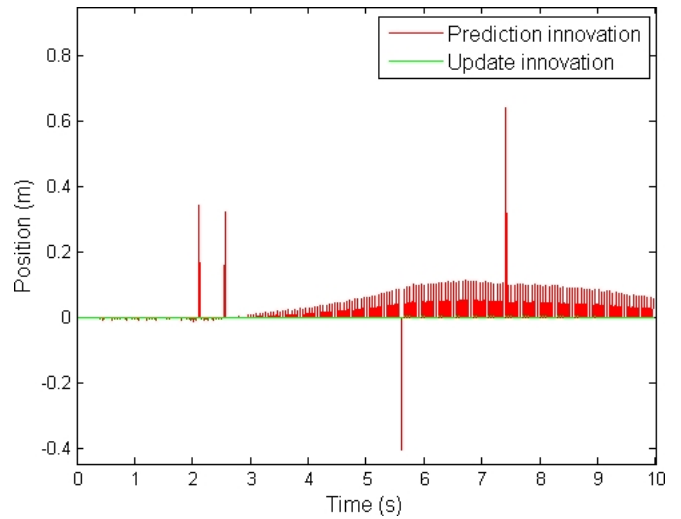


Fig. 6. Position prediction and update innovation sequences.

To avoid increasing difference between process and measurement estimates, modifications are made in prediction equation (16) by adding control vector:

$$\hat{x}_k = A^P \bar{x}_{k-1} + B u_k + w_k, \quad (30)$$

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (31)$$

$$u_k = \begin{bmatrix} \hat{y}_k^m \\ 0 \\ 0 \end{bmatrix}, \quad (32)$$

$$\hat{y}_k^m = \hat{y}_k + \frac{f_{cv}}{f_{sf}} \times \frac{\sum_{i=k-n+1}^k (z_i - z_{i-1})}{(n-1)}, \quad (33)$$

where  $B$  – control matrix;  $u_k$  – control vector  $n = 4$  – four last available measurements from CV. By adding control vector, which consists of modified innovation  $\hat{y}_k^m$  it is ensured that position estimation from SF will not diverge between CV samples. Modified innovation is calculated using previously available  $z_k$  measurements using CV and current estimated position from sensors, until the next available CV position information. Estimated position using Kalman filter for hybrid tracking presented is in Fig. 7.

The main demand for AR hybrid tracking technique is in case of a lost object tracking using CV to provide camera orientation-position information from additional sources – inertial sensors. Summarized results of the accomplished experiments are presented in Table I and Table II.

GT denotes ground truth. In presented orientation-position estimation results KF for hybrid tracking shows the best results: mean and standard deviation is nearest to the ground



truth value. In case of root mean square error in orientation estimation the error is smallest from 90 degrees (an original state in object tracking) and in case of position it is near zero value, which is reliability indication of estimated results.

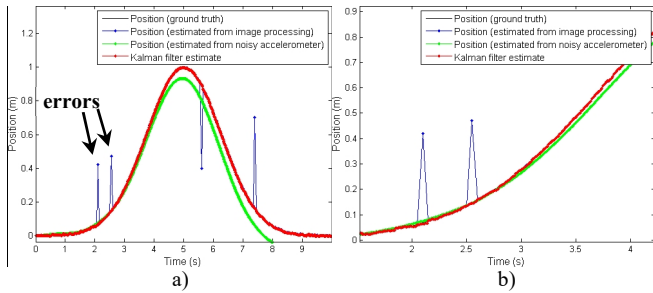


Fig. 7. Hybrid tracking for position estimation.

TABLE I. MEAN ( $\mu$ ), STANDARD DEVIATION ( $\sigma$ ), ROOT MEAN SQUARE ERROR RESULTS FOR ORIENTATION ESTIMATION.

Parameter	Orientation estimation (degrees)			
	GT	SF	CV	KF
$\mu$	90.000	90.300	88.551	90.277
$\sigma$	63.672	63.725	63.843	63.684
RMSE	-	89.727	90.27	89.787

TABLE II. MEAN ( $\mu$ ), STANDARD DEVIATION ( $\sigma$ ), ROOT MEAN SQUARE ERROR RESULTS FOR POSITION ESTIMATION.

Parameter	Position estimation (meters)			
	GT	SF	CV	KF
$\mu$	0.314	0.344	0.319	0.314
$\sigma$	0.351	0.376	0.358	0.352
RMSE	-	0.041	0.062	0.003

Other researches related to the tracking approaches does not provide their experimental results, nor ground truth values in a numeric form to make objective comparisons. As a result, all experiments were accomplished with simulated data. Improvements using Hybrid tracking are presented in Table III.

TABLE III. SUMMARY OF HYBRID TRACKING IMPROVEMENTS.

Tracking technique	Orientation		Position	
	Slow	Fast	Slow	Fast
Based on computer vision	+/-	-	+/-	-
Based on sensor fusion	+	+	+/-	+/-
Based on hybrid	+	+	+	+/-

In case of slower camera motion speed, orientation estimation using CV holds reliable estimates compared to fast motion, when object recognition and tracking in the image is difficult or impossible. In case of CV “+/-” denotes additional conditions: whether the object in the scene is recognized or not. Unavailable or occluded object in the scene is a possibility. SF orientation accuracy is not affected by motion speed. In case of SF “+/-” denotes unreliable position estimation. Kalman filter for hybrid tracking balances between reliable and unreliable estimates from CV and SF estimates. Sign “-” shows that object position and orientation tracking is not possible. It is important to note that computer CV and SF must be implemented on different threads. This allows to avoid system delay and accomplish data processing independently. Even though this research focus on a hybrid tracking technique, solution can be easily adopted for the development of interaction devices.

## VI. CONCLUSIONS

1. Orientation-position estimation approaches were analysed based on sensor fusion and computer vision. Hybrid tracking solution using Kalman filter was proposed that has supplementary properties and eliminates separate tracking technique disadvantages.
2. Orientation and position information were simulated with probable errors that might arise in real life conditions, caused by rapid motion, occlusions and other problems. Hybrid tracking research estimated results were compared with only computer vision and only sensor fusion estimates.
3. Depending on the available information from sources, different approaches to orientation and position estimation problems are proposed incorporating innovation conditions. Proposed solution improves tracking reliability and eliminates delay in computer vision provided orientation-position estimates.
4. Hybrid tracking technique adoption in augmented reality is the main aspect of this research. Orientation-position improved estimates would ensure reliable virtual content representation into real world context with realistic continuous view and interaction.

## REFERENCES

- [1] H. Bay, T. Tuytelaars, L. V. Gool, “SURF: Speeded Up Robust Features”, *Computer Vision and Image Understanding*, vol. 3, no. 110, pp. 346–359, 2008. [Online]. Available: <http://dx.doi.org/10.1016/j.cviu.2007.09.014>
- [2] E. Rosten, R. Porter, T. Drummond, “Faster and better: a machine learning approach to corner detection”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 105–119, 2010. [Online]. Available: <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2008.275>
- [3] J. Polvi, T. Taketomi, G. Yamamoto, A. Dey, C. Sandor, H. Kato, “SLiDAR: A 3D positioning method for SLAM-based handheld augmented reality”, *Computers & Graphics*, vol. 55, pp. 33–43, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.cag.2015.10.013>
- [4] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, D. Schmalstieg, “Real-time detection and tracking for augmented reality on mobile phones”, *IEEE Trans. visualization and computer graphics*, vol. 16, no. 3, pp. 355–368, 2010. [Online]. Available: <http://doi.ieeecomputersociety.org/10.1109/TVCG.2009.99>
- [5] D. Kurz, S. Benhimane, “Handheld Augmented Reality involving gravity measurements”, *Computers & Graphics*, vol. 36, no. 7, pp. 866–883. [Online]. Available: <http://dx.doi.org/10.1016/j.cag.2012.03.038>, 2012
- [6] E. Lukosiuonas, V. Bulbenkiene, A. Andziulis, L. Pasviestis, E. Artemciukas, “An ultrasonic tracking method for augmented reality”, in *Proc. 17th Int. Conf. Information and Software Technologies*, Kaunas, Lithuania, 2011, pp. 170–173.
- [7] J. D. Hol, T. B. Schon, F. Gustafsson, P. J. Slycke, “Sensor fusion for augmented reality”, in *9th Int. Conf. Information Fusion*, Florence, Italy, 2006, pp. 1–6. [Online]. Available: <http://dx.doi.org/10.1109/ICIF.2006.301604>
- [8] S. O. H. Madgwick, “Estimation of IMU and MARG orientation using a gradient descent algorithm”, in *IEEE Int. Conf. Rehabilitation Robotics*, Zurich, Switzerland, 2011, pp. 1–27. [Online]. Available: <http://dx.doi.org/10.1109/ICORR.2011.5975346>
- [9] A. M. Sabatini, “Quaternion-based extended Kalman filter for determining orientation by inertial and magnetic sensing”, *IEEE Trans. Biomedical Engineering*, vol. 7, no. 53, pp. 1346–1356, 2006. [Online]. Available: <http://dx.doi.org/10.1109/TBME.2006.875664>
- [10] P. Daponte, L. De Vito, F. Picariello, M. Riccio, “State of the art and future developments of the Augmented Reality for measurement applications”, *Measurement*, vol. 57, pp. 53–70, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.measurement.2014.07.009>
- [11] G. Bleser, D. Stricker, “Advanced tracking through efficient image processing and visual-inertial sensor fusion”, in *Proc. Virtual Reality 2008 Conf.*, Los Alamitos, California, 2008, pp. 137–144. [Online]. Available: <http://dx.doi.org/10.1109/VR.2008.4480765>