Leopoldo Armesto
Josep Tornero
Dept. of Control Systems Engineering,
Technical University of Valencia
Camino de Vera, s/n 46022, Valencia, Spain
leoaran@isa.upv.es, jtornero@isa.upv.es

Markus Vincze
Automation and Control Institute
Vienna University of Technology
Gusshausstr. 27.29/361 A-1040, Vienna, Austria
vincze@acin.tuwien.ac.at

Fast Ego-motion Estimation with Multi-rate Fusion of Inertial and Vision*

Abstract

This paper presents a tracking system for ego-motion estimation which fuses vision and inertial measurements using EKF and UKF (Extended and Unscented Kalman Filters), where a comparison of their performance has been done. It also considers the multi-rate nature of the sensors: inertial sensing is sampled at a fast sampling frequency while the sampling frequency of vision is lower. the proposed approach uses a constant linear acceleration model and constant angular velocity model based on quaternions, which yields a non-linear model for states and a linear model in measurement equations. Results show that a significant improvement is obtained on the estimation when fusing both measurements with respect to just vision or just inertial measurements. It is also shown that the proposed system can estimate fast-motions even when vision system fails. Moreover, a study of the influence of the noise covariance is also performed, which aims to select their appropriate values at the tuning process. The set-up is an end-effector mounted camera, which allow us to pre-define basic rotational and translational motions for validating results.

KEY WORDS—vision and inertial, multi-rate systems, sensor fusion, tracking

1. Introduction

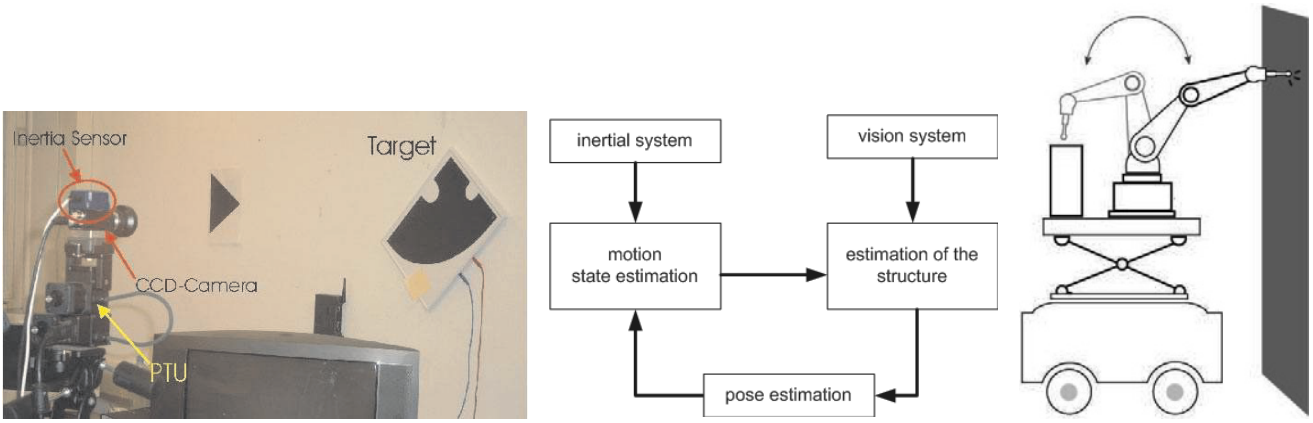
One of the most common techniques for state estimation of non-linear discrete-time dynamic systems is the Extended Kalman filter (EKF) (Dissanayake et al. 2001) or recently, the Unscented Kalman filter (UKF) (Julier and Uhlmann 2002). The Kalman filter gives a robust, optimal, recursive state estimation to fuse redundant information from different sensors. However, both approaches assume that the probability distribution function (*pdf*) of the noise is Gaussian, which is not true for non-linear systems. Other recent filtering methods are Particle Filters (PF) (Gordon et al. 1993; Doucet et al. 2001), where the main advantage is that the *pdf* can be accurately approximated with a large number of particles with a cost $\mathcal{O}(N)$. The most common approach to PF is the Sampling Importance Resampling (SIR) (Smith and Gelfand 1992; Carpenter et al. 1997). Another well-known approach is the Rao–Blackwellized PF (Doucet et al. 2000), which uses a PF for some variables of the state and an EKF filter for other variables.

In mobile robots, inputs and outputs have different sampling rates. Proprioceptive sensors, such as encoders, gyros, accelerometers, might be considered as inputs or outputs, depending of the considered approach, while exteroceptive sensors, such as laser, sonar rangers, vision systems, etc., are usually considered as outputs. Proprioceptive sensors are typically sampled several times faster than exteroceptive sensors, see for instance (Armesto and Tornero 2004; Huster and Rock 2003; Armesto et al. 2004). This is a problem that arises from inherent technological limitations of each type of sensor, communication channels, processing cost, etc. A typical solution to overcome this problem is to increase the overall sampling period to the slowest one. However, it is well known that this

The International Journal of Robotics Research
Vol. 26, No. 6, June 2007, pp. 577–589
DOI: 10.1177/0278364907079283

Figures 1, 4, 6–9, 11, 12 appear in color online: http://ijr.sagepub.com

* This work has been supported by the Spanish Goverment (Ministerio de Ciencia y Tecnologia), research project PDI2000-0362-P4-05 and BIA2005-09377-C03-02 the Austrian Science Foundation (FWF) under grant P15748. The paper was received on 10/04/2006, revised on 23/01/2007 and accepted on 03/03/2007.



(a) Pan-Tilt head with monocular-camera and IMU Tracking System.

(b) Tracking System Diagram.

(c) Schematic idea of RESTAURO Research Project.

Fig. 1. Applications to the tracking system for ego-motion estimation.

approach may decrease the overall system performance since high frequency dynamics are missed due to the temporal discretization as a consequence of the Nyquist sampling constraint.

Multi-rate systems have been extensively treated in the last four decades and it is possible to find many contributions dealing with modelling (Albertos 1990; Khargonekar et al. 1985; Kranc 1957; Tornero 1985) and analysis (Goodwin and Feuer 1992) as well as control design (Tornero et al. 1999) of periodic sampled-data systems. One of the most relevant modelling techniques is the Lifting Technique (Khargonekar et al. 1985), where an isomorphism between a linear periodic system and an enlarged linear-time invariant (LTI) system is defined via the lifting operator. Another interesting point of view for modelling multi-rate systems is the one provided by Tornero (1985) and Longhi (1994), where two periodic matrices relate inputs and outputs according to the multi-rate sampling pattern. The main advantage of this approach with respect to the lifting technique is that it is not restricted to linear systems and it is implemented at the fastest sampling rate, and therefore it is much more appropriate for "real-time" systems.

This paper investigates a new, generic approach to multi-rate tracking combining vision and inertial sensor data. The fusion of vision and inertial measurements provides complementary characteristics: visual sensing is very accurate at low velocities while inertial sensors can track fast motions but suffer from drift, particularly at low velocities. The set-up for this application is an end-effector mounted camera together with an inertial sensor based on accelerometers and gyroscopes.

One of the motivations of this tracking system is to estimate arbitrary motions of mobile robots or people. The application is to use a monocular-camera vision system and an inertial measurement unit (IMU) mounted on a pan–tilt unit (PTU), as shown in Figure 1(a). In Figure 1(a), the system diagram of the tracking system is shown in 1(b). In addition, this tracking system will accurately estimate the pose of the end-effector of a manipulator, used for restoration tasks of old-building facades. This robot will be mounted on a mobile base and a lifting structure as shown in Figure 1(c). The purpose of this mechanical system is to extend the work area of the manipulator, although it has the disadvantage of suffering from perturbations and/or oscillations.

This paper presents a precise model description for pose estimation where jerks and angular accelerations are treated as noise, including the effects of centripetal accelerations. With respect to our previous work (Armesto et al. 2004), this paper is focused on the study of estimation of different tracking velocities in rotational and translational movements.

As a main contribution, fusion of vision and inertial measurements is performed with a generic multi-rate EKF (MR-EKF) and multi-rate UKF (MR-UKF), in order to deal with data at different sampling rates. This approach improves the overall performance with respect to single-rate methods. Moreover, this fusion concept is valid for other sensors such as laser rangers and encoders in mobile robot localisation and map building (Armesto and Tornero 2004; Armesto et al. 2007).

In the paper, the influence of uncertainties associated to both type of sensors has been studied. Results show that the combination of vision and inertial data gives better estimation in a wide variety of situations (slow and fast motions).

Results have been obtained offline using Matlab, although real-time results have also been obtained on an implemented version in Labview. Data, Matlab code and Videos can be found in 2 at the Appendix.

1.1. Related Work

Fusion of inertial and vision data is needed in many applications (Rehbinder and Ghosh 2001; Alves et al. 2003; Huster and Rock 2003; Abuhadrous et al. 2003; Chroust and Vincze 2003; Panerai et al. 2000), specially for designing new "intelligent" cameras with integrated low-cost accelerometers and gyroscopes (MEMS). In Rehbinder and Ghosh (2001), inertial and vision data (with a delay), acting at different sampling rates, are used for the rotation estimation with an observer. For modelling and calibration of vision and inertial sensors see Alves et al. (2003), where an algorithm is presented to estimate the internal calibration of the inertial sensor and also to estimate the relative orientation between camera and inertial system. An UKF approach of data fusion with vision and inertial sensors is described in Huster and Rock (2003). In Lobo and Dias (1998), a pose estimation algorithm is proposed to integrate measurements from a stereo camera system and an inertial sensor for mobile robots. Due to the restricted motion of the robot (only in a plane) the estimation routine is simplified significantly as only two positions and one angle have to be estimated to determine the pose of the robot. Sandini and co-workers investigated a stabilization gaze algorithm in order to obtain "steady-state" images (Panerai et al. 2000) using vision and inertial data. Their results show that stabilized images improve the reactivity to changes of the environment.

In Jekeli (2001) and Grewal et al. (2001) a thorough revision of pose estimation is given based on Inertial/GPS fusion. These systems take into account several aspects regarding with Earth geodetic properties such as earth rotation (Coriolis effect), Earth elliptic shape, complex gravity models, etc. Inertial/GPS fusion has many similarities with Inertial/Vision fusion, although they might be applied in different environments. In the present paper, we neglect Earth properties such those used in Jekeli (2001) and Grewal et al. (2001), since the scope of our application is limited to smaller areas where the Earth plain surface assumption holds.

Huster and Rock (2001, 2003) developed an algorithm to estimate the pose of autonomous underwater vehicles. The state description combines the position, velocity and acceleration together with quaternion representation of the orientation. The state also includes an estimation of the acceleration bias. They analyzed an EKF, a two-step estimator and a UKF, which has been shown to be the most appropriate filter for these vehicles. Their conclusion was that better estimation results were obtained if the non-linearity is in the states only, with linear measurement equations. In Gurfill and Kasdin (2002) similar conclusions are given. Based on these conclusions, our approach uses a non-linear model for states and a linear model for the measurement equations.

In Rehbinder and Ghosh (2001), a nonlinear estimator for the fusion of inertial and vision systems based on geometric properties has been developed. An architecture to estimate the ego-motion and the structure of an environment using inertial and vision sensors is also presented in Chai et al. (2002). The dynamic description of the system combines an acceleration model in the linear motion and a velocity model in the orientation. The measurement model directly uses image coordinates, which results in a nonlinear measurement model besides the nonlinear dynamic model.

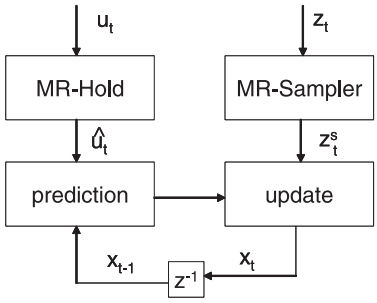


Fig. 2. General multi-rate filtering scheme.

Pose estimation based on inertial sensors using quaternions is treated in detail in Goddard and Abidi (1998) where dynamic equations are derived and linearised. In Ude (1999), an approach is developed based on exponential mapping for predicting quaternions depending on angular velocities, in a mathematical way. Another way to predict quaternions is the use of a so-called *quaternion velocity* (Chou 1992), which has two disadvantages, first it leads to a nonlinear measurement equation, which results in poorer estimation results (Gurfill and Kasdin 2002) and secondly it is only an approximation valid for low angular velocities. The approach of Ude (1999) has been extended in this paper, taking into account angular accelerations in addition to angular velocities.

2. Multi-rate Filtering

In this paper, a multi-rate filtering structure is used based on two different sampling frequency interfaces: multi-rate holds (MR-Holds) and samplers (MR-Samplers), as schematized in Figure 2. These MR-Interfaces allow us to implement conventional Prediction and Update steps at a the fastest sampling rate. Multi-rate holds are used to interface inputs of the system at the prediction step, while multi-rate samplers interface outputs at the update step. It is interesting to note that the Prediction and Update steps run at the fastest frequency, while each signal is sampled at its own sampling rate.

In the context of mobile robot pose estimation, signals from proprioceptive sensors such as encoders, accelerometers, gyroscopes, etc. might be considered as inputs ($\mathbf{u}_t$) or outputs ($\mathbf{z}_t$), while signals from exteroceptive sensors such as laser and sonar rangers, GPS or vision systems are always considered as outputs ($\mathbf{z}_t$). If all measurements are considered as outputs, only MR-Samplers are required, where such considerations depend on associated dynamic models.

Table 1. Multi-rate HOH primitive functions

Type	Function
Interpolation (Lagrange)	$\hat{\mathbf{u}}(t) = \sum_{l=0}^{n} \left[\prod_{\substack{q=0 \\ q \neq l}}^{n} \left[\frac{t - t_{j-q}}{t_{j-l} - t_{j-q}} \right] \mathbf{u}(t_{j-l}) \right]$
Approximation (Bezier)	$\hat{\mathbf{u}}(t) = \sum_{l=0}^{n} \left[\frac{n!}{l!(n-l)!} \left(1 + \frac{t - t_j}{t_j - t_{j-n}} \right)^{n-l} \left(\frac{t - t_j}{t_j - t_{j-n}} \right)^{l} \mathbf{u}(t_{j-l}) \right]$
Approximation (Taylor)	$\hat{\mathbf{u}}(t) = \sum_{l=0}^{n} \frac{(t - t_j)^l}{l!} \mathbf{u}^{l)}(t_j)$

2.1. Multi-rate Holds

The mathematical background of multi-rate holds and samplers is described in our previous contributions (Tornero et al. 1999; Tornero and Tomizuka 2000, 2002). A multi-rate hold is a hybrid *device* for generating, from a sequence of inputs sampled at a slow sampling rate, a continuous signal which is discretized at a high sampling rate. In these contributions, multi-rate zero, first and second order holds (MR-ZOH, MR-FOH and MR-SOH) were presented. Later on, in Armesto and Tornero (2003), a wide variety of holds were obtained based on primitive functions. The idea behind Armesto and Tornero (2003) is to generate an extrapolated continuous signal based on previous input samplings $\{\mathbf{u}(t_j), \mathbf{u}(t_{j-1}), \ldots, \mathbf{u}(t_{j-n})\}$, uniformly distributed or not, where j represent each time instant an input is sampled.

$$\hat{\mathbf{u}}(t) = \sum_{l=0}^{n} \mathbf{f}_l(t, \mathbf{u}(t_{j-l}), t_{j-l}) \tag{1}$$

where $\mathbf{u}(t_j)$ denotes an input that has been sampled at time instant t_j, where $t_{j-n} < \ldots < t_j < t$. The primitive function $\mathbf{f}_l(t)$ generates the continuous signal $\hat{\mathbf{u}}(t)$, which afterwards is discretized at any desired sampling rate T to provide $\hat{\mathbf{u}}_k$, with $t = kT$. In order to implement the multi-rate hold, a shift register $\mathcal{U}_j = \{\mathbf{u}_j, \ldots, \mathbf{u}_{j-n}\}$ is required to maintain the history of the signal. Algorithm 1 implements the (discrete-time) multi-rate hold based on a general primitive function, where a constant base period T is assumed. If an input arrives during sampling period T, lines 3 to 5 are executed, otherwise lines 7 to 10 perform an extrapolation process based on the registered history of the signal $\mathcal{U}_j$, while j represents the time instant of the most recent sample added to $\mathcal{U}_j$. Asynchronous holds, with variable sampling periods, can also be found in Armesto (2007).

Table 1 summarizes some primitive functions that that can be used in multi-rate holds. In particular, for Taylor holds, derivatives are computed using the backward approximation:

$$\mathbf{u}^{n)}(t_j) = \frac{\mathbf{u}^{n-1)}(t_j) - \mathbf{u}^{n-1)}(t_{j-1})}{t_j - t_{j-1}} \tag{2}$$

where $n)$ means the nth derivative.

Algorithm 1 MR-Hold

```
1  MR-Hold(u_k, U_j, k, j)
2  if u_k is sampled then
3     shift out u_{j-n} and shift in u_k from U_j;
4     û_k = u_k;
5     j = k;
6  else
7     û_k = 0;
8     for l = 0 to n do
9        retrieve u_{j-l} from U_j;
10       û_k = û_k + f_l(kT, u_{j-l}, (j - l)T);
11    end
12 end
13 return û_k, U_j and j;
```

2.2. Multi-rate Samplers

A multi-rate sampler is used to interface outputs of a system sampled at different sampling rates. As a result, it generates a size-varying signal at a fast sampling rate that combines measurements at different samplings. For a given time-instant, the multi-rate sampler appends to the measurement vector $\mathbf{y}_k^s$ the measurement $\mathbf{y}_{i,k}$ if and only if the sensor has performed a valid acquisition, thus $\mathbf{y}_{i,k} \subset \mathbf{y}_k^s$, if $\mathbf{y}_{i,k}$ is available (or sampled), where i denotes the ith measurement.

Although this concept is not new, it is important in sensor fusion techniques to coherently integrate measurements at different sampling rates into the estimation.

Take Figure 3 as an example of periodic sampling with $N = 6$, where N is the periodicity ratio within the frame-period, the sampling period where signals are periodically repeated. According to this, the resulting size-varying vector $\mathbf{y}_k^s$ is:

$$\begin{aligned} \mathbf{y}_{jN}^s &= [y_{1,jN}, y_{2,jN}]^T, \quad \mathbf{y}_{jN+1}^s = [\emptyset], \\ \mathbf{y}_{jN+2}^s &= [y_{2,jN+2}], \quad \mathbf{y}_{jN+3}^s = [y_{1,jN+3}], \\ \mathbf{y}_{jN+4}^s &= [y_{2,jN+4}], \quad \mathbf{y}_{jN+5}^s = [\emptyset]. \end{aligned} \tag{3}$$

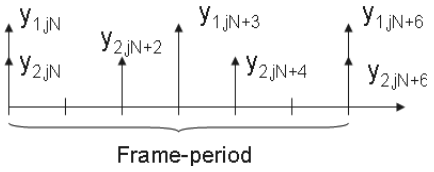


Fig. 3. Sampling example.

2.3. Multi-rate Extended Kalman Filter

A general discrete-time stochastic non-linear system can be described as:

$$\mathbf{x}_k = \mathbf{f}\left[\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}\right] \quad (4)$$

$$\mathbf{y}_k = \mathbf{h}\left[\mathbf{x}_k\right] + \mathbf{v}_k. \quad (5)$$

Equations (4) and (5) describe the dynamic behaviour of the system, with $\mathbf{f}[\bullet]$ the non-linear vector-valued function relating the system dynamics, and $\mathbf{h}[\bullet]$ the non-linear output measurement equations. The state vector $\mathbf{x}_k$, the system noise $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$ are assumed to be variables with Gaussian distribution, with covariances $\mathbf{P}_k$, $\mathbf{Q}_k$ and $\mathbf{R}_k$ respectively.

The multi-rate Extended Kalman filter can be implemented as described in Algorithm 2. First the multi-rate hold estimates the control input (line 2), if available. Second, the multi-rate sampler is implemented so that it generates a size-varying output vector $\mathbf{y}_k^s$, output equations $\mathbf{h}^s$ and output covariance $\mathbf{R}_k^s$ (lines 3 to 10, where sub-index i represent the element within a vector/matrix). Lines 11 to 15 implement the conventional EKF prediction equations, while EKF update (lines 21 and 22) is only produced if any output has been sampled (length of $\mathbf{y}_k^s$ is greater than zero). Otherwise, the prediction is taken as the best guess for the next iteration.

Algorithm 2 Multi-rate Extended Kalman Filter.

1 **MR-EKF**$\left(\mathbf{x}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \mathbf{Q}_{k-1}, \mathbf{R}_k, \mathbf{u}_{k-1}, \mathbf{y}_k, \mathcal{U}_j, k, j\right)$
2 $[\hat{\mathbf{u}}_{k-1}, \mathcal{U}_j, j] =$ MR-Hold $(\mathbf{u}_{k-1}, \mathcal{U}_j, k-1, j)$;
// Multi-rate output sampler
3 $\mathbf{y}_k^s = \emptyset$; $\mathbf{h}^s = \emptyset$; $\mathbf{R}_k^s = \emptyset$;
4 **for** $i = 1$ **to** length $(\mathbf{y}_k)$ **do**
5 **if** sensor i is sampled **then**
6 append in rows $\mathbf{y}_{i,k}$ to $\mathbf{y}_k^s$;
7 append in rows $\mathbf{h}_i[\bullet]$ to $\mathbf{h}^s[\bullet]$;
8 append in rows and columns $\mathbf{R}_{i,k}$ to $\mathbf{R}_k^s$;
9 **end**
10 **end**
11 $\nabla\mathbf{f}_x = \left.\frac{\partial \mathbf{f}[\bullet]}{\partial \mathbf{x}_{k-1}}\right|_{\hat{\mathbf{x}}_{k-1|k-1}}$;
12 $\nabla\mathbf{f}_w = \left.\frac{\partial \mathbf{f}[\bullet]}{\partial \mathbf{w}_{k-1}}\right|_{\hat{\mathbf{x}}_{k-1|k-1}}$;
13 $\nabla\mathbf{h}_x^s = \left.\frac{\partial \mathbf{h}^s[\bullet]}{\partial \mathbf{x}_k}\right|_{\hat{\mathbf{x}}_{k|k-1}}$;
14 $\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}\left[\hat{\mathbf{x}}_{k-1|k-1}, \hat{\mathbf{u}}_{k-1}, \mathbf{0}\right]$;
15 $\mathbf{P}_{k|k-1} = \nabla\mathbf{f}_x \mathbf{P}_{k-1|k-1} \nabla\mathbf{f}_x^T + \nabla\mathbf{f}_w \mathbf{Q}_{k-1} \nabla\mathbf{f}_w^T$;
16 **if** length $(\mathbf{y}_k^s) > 0$ **then**
17 $\mathbf{K}_k = \mathbf{P}_{k|k-1}(\nabla\mathbf{h}_x^s)^T \left[\nabla\mathbf{h}_x^s \mathbf{P}_{k|k-1} (\nabla\mathbf{h}_x^s)^T + \mathbf{R}_k^s\right]^{-1}$;
18 $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k^s - \mathbf{h}^s[\hat{\mathbf{x}}_{k|k-1}])$;
19 $\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \nabla\mathbf{h}_x^s \mathbf{P}_{k|k-1}$;
20 **else**
21 $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1}$;
22 $\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1}$;
23 **end**
24 **return** $\mathbf{x}_{k|k}, \mathbf{P}_{k|k}, \mathcal{U}_j$ and j;

2.4. Multi-rate Unscented Kalman Filter

In a similar way, a Multi-rate Unscented Kalman filter can be implemented as described in Algorithm 3. First the multi-rate hold estimates the control input (line 2) and the multi-rate sampler is implemented as before (lines 3 to 10). The Unscented Transform is implemented in lines 11–22, prediction equations of the Unscented Kalman Filter are implemented in lines 23 to 30 and update equations are implemented in lines 32–43. Again, if no measurement has been received, the prediction is taken as the best guess for the next iteration (lines 45 and 46).

3. Motion model

The state of the tracking system is composed of position and orientation variables. Position is described with Cartesian positions $\mathbf{p}_k$ together with their velocities $\mathbf{v}_k$ and accelerations $\mathbf{a}_k$. The orientation is represented with quaternions $\mathbf{q}_k$ and angular velocities $\boldsymbol{\omega}_k$. Previous results (Huster and Rock 2001; Chroust and Vincze 2003) have shown an improvement if the biases of the acceleration measurements $\mathbf{b}_k$ are included and estimated on-line. The output vector $\mathbf{y}_k$ is formed with the measured accelerations $\mathbf{a}_k^m$ and angular velocities $\boldsymbol{\omega}_k^m$ from the inertial sensor and the Cartesian positions $\mathbf{p}_k^m$ and quaternions $\mathbf{q}_k^m$ from the vision system, which have been obtained after the image processing procedure. Jerks $\mathbf{j}_k$, in the Cartesian coordinates, angular accelerations $\boldsymbol{\alpha}_k$ and velocity biases $\mathbf{b}_k'$ are considered as the system noise. As usual, the output vector has associated a measurement noise vector $\boldsymbol{\upsilon}_k$. In this case, the system has no inputs, since command movements are assumed to be unknown.

$$\mathbf{x}_k = [\mathbf{p}^T\ \mathbf{v}^T\ \mathbf{a}^T\ \mathbf{b}^T\ \mathbf{q}^T\ \vec{\omega}^T]_k^T \quad (6)$$

Algorithm 3 Multi-rate Unscented Kalman Filter.

MR-UKF$(\mathbf{x}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \mathbf{Q}_{k-1}, \mathbf{R}_k, \mathbf{u}_{k-1}, \mathbf{y}_k, \mathcal{U}_j, k, j)$
$[\hat{\mathbf{u}}_{k-1}, \mathcal{U}_j, j] = \text{MR-Hold}(\mathbf{u}_{k-1}, \mathcal{U}_j, k-1, j)$;
// Multi-rate output sampler
$\mathbf{y}_k^s = \emptyset$; $\mathbf{h}^s = \emptyset$; $\mathbf{R}_k^s = \emptyset$;
for $i = 1$ **to** length $(\mathbf{y}_k)$ **do**
  if sensor i is sampled **then**
    append in rows $\mathbf{y}_{i,k}$ to $\mathbf{y}_k^s$;
    append in rows $\mathbf{h}_i$ [•] to $\mathbf{h}^s$ [•];
    append in rows and columns $\mathbf{R}_{i,k}$ to $\mathbf{R}_k^s$;
  end
end
$\mathbf{x}^a = \begin{bmatrix} \hat{\mathbf{x}}_{k-1|k-1}^T & \mathbf{0} & \mathbf{0} \end{bmatrix}^T$;
$\mathbf{P}^a = \begin{pmatrix} \mathbf{P}_{k-1|k-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_k^s \end{pmatrix}$;
$\mathcal{P} = \sqrt{\mathbf{P}^a}$; $\mathcal{X}_0^a = \mathbf{x}^a$; $W_0 = \kappa/(n+\kappa)$;
for $i = 1$ **to** $\tilde{n}$ **do**
  $\mathcal{X}_i^a = \mathbf{x}^a + \mathcal{P}_i$; $W_i = 1/(2(n+\kappa))$;
  $\mathcal{X}_{i+n}^a = \mathbf{x}^a - \mathcal{P}_i$; $W_{i+n} = 1/(2(n+\kappa))$;
end
for $i = 0$ **to** $2\tilde{n}$ **do**
  $\mathcal{X}_{i,k-1|k-1} = \mathcal{X}_{i,1:n}^a$;
  $\mathcal{W}_{i,k-1} = \mathcal{X}_{i,n+1:n+g}^a$;
  $\mathcal{V}_{i,k}^s = \mathcal{X}_{i,n+g+1:n+g+r}^a$;
end
$\hat{\mathbf{x}}_{k|k-1} = \mathbf{0}$; $\mathbf{P}_{k|k-1} = \mathbf{0}$
for $i = 0$ **to** $2n$ **do**
  $\mathcal{X}_{i,k|k-1} = \mathbf{f}\left[\mathcal{X}_{i,k-1|k-1}, \mathbf{u}_{k-1}, \mathcal{W}_{i,k-1}\right]$;
  $\hat{\mathbf{x}}_{k|k-1} = \hat{\mathbf{x}}_{k|k-1} + W_i \mathcal{X}_{i,k|k-1}$;
end
for $i = 0$ **to** $2n$ **do**
  $\mathbf{P}_{k|k-1} = \mathbf{P}_{k|k-1} + W_i \left[\mathcal{X}_{i,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\right]\left[\mathcal{X}_{i,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\right]^T$
end
if length $(\mathbf{y}_k^s) > 0$ **then**
  $\hat{\mathbf{y}}_{k|k-1}^s = \mathbf{0}$; $\mathbf{P}_{k|k-1}^{yy} = \mathbf{0}$; $\mathbf{P}_{k|k-1}^{xy} = \mathbf{0}$;
  for $i = 0$ **to** $2n$ **do**
    $\mathcal{Y}_{i,k|k-1}^s = \mathbf{h}^s[\mathcal{X}_{i,k|k-1}] + \mathcal{V}_{i,k}^s$;
    $\hat{\mathbf{y}}_{k|k-1}^s = \hat{\mathbf{y}}_{k|k-1}^s + W_i \mathcal{Y}_{i,k|k-1}^s$;
  end
  for $i = 0$ **to** $2n$ **do**
    $\mathbf{P}_{k|k-1}^{yy} = \mathbf{P}_{k|k-1}^{yy} + W_i\left[\mathcal{Y}_{i,k|k-1}^s - \hat{\mathbf{y}}_{k|k-1}^s\right]\left[\mathcal{Y}_{i,k|k-1}^s - \hat{\mathbf{y}}_{k|k-1}^s\right]^T$;
    $\mathbf{P}_{k|k-1}^{xy} = \mathbf{P}_{k|k-1}^{xy} + W_i\left[\mathcal{X}_{i,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\right]\left[\mathcal{Y}_{i,k|k-1}^s - \hat{\mathbf{y}}_{k|k-1}^s\right]^T$;
  end
  $\mathbf{K}_k = \mathbf{P}_{k|k-1}^{xy}(\mathbf{P}_{k|k-1}^{yy})^{-1}$;
  $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k^s - \hat{\mathbf{y}}_{k|k-1}^s)$;
  $\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{k|k-1}^{yy} \mathbf{K}_k^T$;
else
  $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1}$;
  $\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1}$;
end
return $\mathbf{x}_{k|k}, \mathbf{P}_{k|k}, \mathcal{U}_j$ and j;

$$\mathbf{y}_k = [(\mathbf{a}^m)^T\ (\boldsymbol{\omega}^m)^T\ (\mathbf{p}^m)^T\ (\mathbf{q}^m)^T]_k^T \tag{7}$$

$$\mathbf{w}_k = [\mathbf{j}^T\ \boldsymbol{\alpha}^T\ \mathbf{b}'^T]_k^T \tag{8}$$

$$\boldsymbol{\upsilon}_k = [\boldsymbol{\upsilon}_{a^m}^T\ \boldsymbol{\upsilon}_{\omega^m}^T\ \boldsymbol{\upsilon}_{p^m}^T\ \boldsymbol{\upsilon}_{q^m}^T]_k^T. \tag{9}$$

The acceleration of a system subject to a rotational and translational motion is an inertial frame is $\mathbf{a} = \mathbf{a}^t + \boldsymbol{\omega} \times \mathbf{v}$, where $\mathbf{a}^t$ is the tangential acceleration. The derivative of the acceleration is:

$$\begin{aligned} \mathbf{a}' &= \frac{d\mathbf{a}}{dt} = \frac{d\mathbf{a}^t}{dt} + \frac{d\boldsymbol{\omega}}{dt} \times \mathbf{v} + \boldsymbol{\omega} \times \frac{d\mathbf{v}}{dt} \\ &= \mathbf{j} + \boldsymbol{\alpha} \times \mathbf{v} + \boldsymbol{\omega} \times \mathbf{a}. \end{aligned} \tag{10}$$

Thus, the discrete-time dynamic equations for the acceleration, velocity and position are:

$$\mathbf{p}_{k+1} = \mathbf{p}_k + T \cdot \mathbf{v}_k + \frac{T^2}{2} \cdot \mathbf{a}_k + \frac{T^3}{6} \cdot \mathbf{a}'_k \tag{11}$$

$$\mathbf{v}_{k+1} = \mathbf{v}_k + T \cdot \mathbf{a}_k + \frac{T^2}{2} \cdot \mathbf{a}'_k \tag{12}$$

$$\begin{aligned} \mathbf{a}_{k+1} &= \mathbf{a}_k + T \cdot \mathbf{a}'_k \\ &= \mathbf{a}_k + T \cdot (\mathbf{j}_k + \boldsymbol{\alpha}_k \times \mathbf{v}_k + \boldsymbol{\omega}_k \times \mathbf{a}_k) \end{aligned} \tag{13}$$

where T is the sampling period.

Biases are supposed to be static, but unknown, thus their dynamic equations are assumed to be:

$$\mathbf{b}_{k+1} = \mathbf{b}_k + T \cdot \mathbf{b}'_k. \tag{14}$$

For the representation of the orientation we use quaternions, see Chou (1992) for definition and calculation rules. The dynamic description of a quaternion is given by Ude (1999):

$$\mathbf{q}_{k+1} = \mathbf{q}_{k+1|k} \otimes \mathbf{q}_k = \exp\left(\frac{\Delta\boldsymbol{\theta}_k}{2}\right) \otimes \mathbf{q}_k \tag{15}$$

where,

$$\exp\left(\frac{\Delta\boldsymbol{\theta}_k}{2}\right) = \begin{cases} \begin{bmatrix} \cos\left(\left\|\frac{\Delta\boldsymbol{\theta}_k}{2}\right\|\right) \\ \sin\left(\left\|\frac{\Delta\boldsymbol{\theta}_k}{2}\right\|\right)\frac{\Delta\boldsymbol{\theta}_k}{\|\Delta\boldsymbol{\theta}_k\|} \end{bmatrix}, & \|\Delta\boldsymbol{\theta}_k\| \neq 0 \\ \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^T, & \|\Delta\boldsymbol{\theta}_k\| = 0 \end{cases} \quad (16)$$

and $\otimes$ is the quaternion multiplication (Chou 1992), $\Delta\boldsymbol{\theta}_k$ is the increment of the rotation and $\mathbf{q}_k = [q_0\ q_1\ q_2\ q_3]_k^T$. The norm of both quaternion $\mathbf{q}_{k+1|k}$ and $\mathbf{q}_k$ is 1 and therefore, the norm of the predicted quaternion is also 1.

Now, consider that $\Delta\boldsymbol{\theta}_k$ is:

$$\Delta\boldsymbol{\theta}_k \approx \boldsymbol{\omega}_k \cdot T + \frac{1}{2}\cdot\boldsymbol{\alpha}_k \cdot T^2 \quad (17)$$

which is valid for low sampling periods.

The prediction of quaternions if $\Delta\boldsymbol{\theta}_k = \mathbf{0}$ then $\mathbf{q}_{k+1} = \mathbf{q}_k$, otherwise can be expressed as,

$$\mathbf{q}_{k+1} = \begin{bmatrix} \cos\left(\left\|\boldsymbol{\omega}_k + \frac{T}{2}\boldsymbol{\alpha}_k\right\|\frac{T}{2}\right) \\ \frac{\sin\left(\left\|\boldsymbol{\omega}_k+\frac{T}{2}\boldsymbol{\alpha}_k\right\|\frac{T}{2}\right)}{\left\|\boldsymbol{\omega}_k+\frac{T}{2}\boldsymbol{\alpha}_k\right\|}\cdot\left(\boldsymbol{\omega}_k + \frac{T}{2}\boldsymbol{\alpha}_k\right) \end{bmatrix} \otimes \mathbf{q}_k \quad (18)$$

$$\mathbf{q}_{k+1} = \mathbf{q}_k \cdot \cos\left(\left\|\boldsymbol{\omega}_k + \frac{T}{2}\boldsymbol{\alpha}_k\right\|\frac{T}{2}\right) + \Omega_k \cdot \left(\boldsymbol{\omega}_k + \frac{T}{2}\boldsymbol{\alpha}_k\right)\cdot\frac{\sin\left(\left\|\boldsymbol{\omega}_k + \frac{T}{2}\boldsymbol{\alpha}_k\right\|\frac{T}{2}\right)}{\left\|\boldsymbol{\omega}_k + \frac{T}{2}\boldsymbol{\alpha}_k\right\|}, \quad (19)$$

where,

$$\Omega_k = \begin{bmatrix} -q_1 & -q_2 & -q_3 \\ q_0 & q_3 & -q_2 \\ -q_3 & q_0 & q_1 \\ q_2 & -q_1 & q_0 \end{bmatrix}_k. \quad (20)$$

The prediction of the angular velocity $\boldsymbol{\omega}_k$ is simply performed as

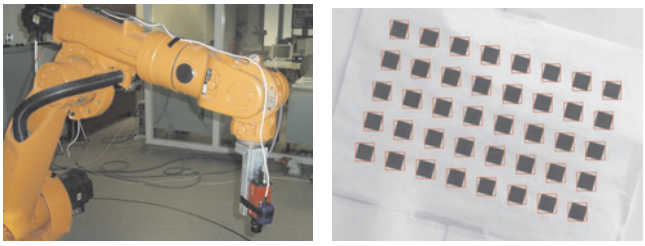
$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k + T\cdot\boldsymbol{\alpha}_k. \quad (21)$$

Finally, the output equations (5), are linear:

$$\mathbf{y}_k = \mathbf{h}[\mathbf{x}_k] + \mathbf{v}_k = \mathbf{H}\cdot\mathbf{x}_k + \mathbf{v}_k \quad (22)$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times4} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times4} \\ \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times4} \\ \mathbf{0}_{4\times3} & \mathbf{0}_{4\times3} & \mathbf{0}_{4\times3} & \mathbf{0}_{4\times3} & \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} \end{bmatrix} \quad (23)$$

where $\mathbf{a}_k^m$ and $\boldsymbol{\omega}_k^m$ in $\mathbf{y}_k$ have been properly rotated based on the estimated orientation to avoid non-linearities on the output equation. A further important issue is the influence of gravity on the measured accelerations. At present, the implemented solution rotates the gravity vector $\mathbf{g} = [0\ 0\ -g]^T$ around the estimated orientation and subtracts this from the measured acceleration.



(a) Vision and inertial sensors mounted on an industrial robot. (b) Image processing example with a grid pattern.

Fig. 4. Experiment set-up.

4. Experimental Results

The proposed tracking system is tested with a predefined set of rotational and translational movements at different speeds. In particular, rotational movements consist of a sequence of turns in roll θ, pitch γ and yaw φ angles:

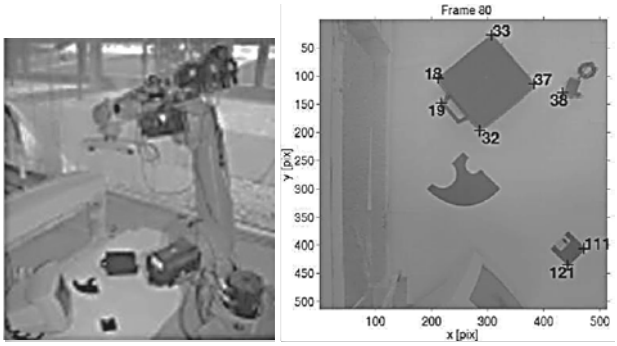
- $\mathbf{R}_z(\varphi)$, $\mathbf{R}_z(-2\varphi)$ and $\mathbf{R}_z(\varphi)$.
- $\mathbf{R}_x(-\theta)$, $\mathbf{R}_x(2\theta)$ and $\mathbf{R}_x(-\theta)$.
- $\mathbf{R}_y(-\gamma)$, $\mathbf{R}_y(2\gamma)$ and $\mathbf{R}_y(-\gamma)$.

Similarly, translational movements consist of:

- $\mathbf{T}_x(-x)$, $\mathbf{T}_x(2x)$ and $\mathbf{T}_x(-x)$.
- $\mathbf{T}_y(y)$, $\mathbf{T}_y(-2y)$ and $\mathbf{T}_y(y)$.

Four different speeds have been tested using different velocities (percentages of the maximum velocity of KUKA KR15/2 robot): 10%, 30%, 50% and 75% where the system has been tested. The system is composed of a firewire camera (Allied Dolphin F145C) with a sampling period of $T_V = 80$ms and an inertial sensor (XSens-MT9B) with a sampling period of $T_I = 10$ms, as shown in Figure 4(a). See Extension 3 in the Appendix.

Pose estimations are computed by detecting squared features on a grid image, as shown in Figure 4(b), using the IMAQ Toolkit of Labview and the Zhang (1998) method. This simplified approach allows us to compute the 3D pose independently from previous estimations, since positions of grid points are assumed to be known. In addition, this approach allows us to focus the research on analyzing the effects of the multi-rate



(a) Set-up (b) Image processing example.

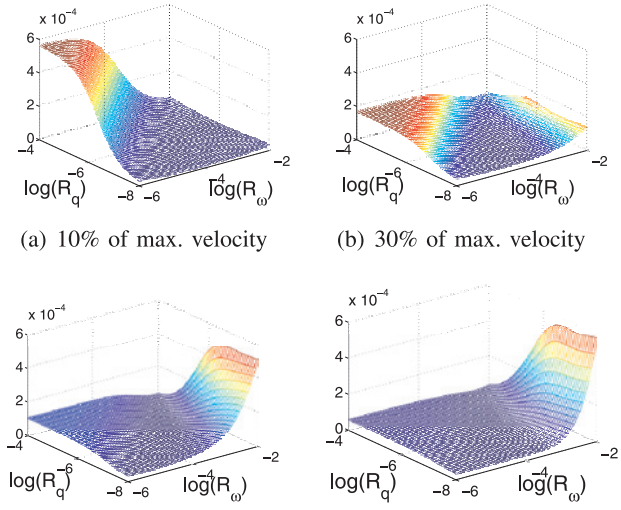
Fig. 5. Previous experiment set-up.

fusion as well as the influence of the covariance $\mathbf{R}_k$ on the estimation. Other well-known approaches of Computer Vision techniques have been used previously in our application based on artificial landmarks and common office objects (Vincze et al. 2001; Bradski 2000; Gemeiner et al. 2006) as shown in Figures 5(a) and 5(b).

A performance index is used to quantify the estimation error, that is the mean quadratic error between estimated positions and orientations and movement references, known completely known for this set-up:

$$J_p = \frac{1}{N}\sum_{i=0}^{N}(\mathbf{p}_{ref} - \mathbf{p})^2,$$
$$J_q = \frac{1}{N}\sum_{i=0}^{N}(\mathbf{q}_{ref} - \mathbf{q})^2. \quad (24)$$

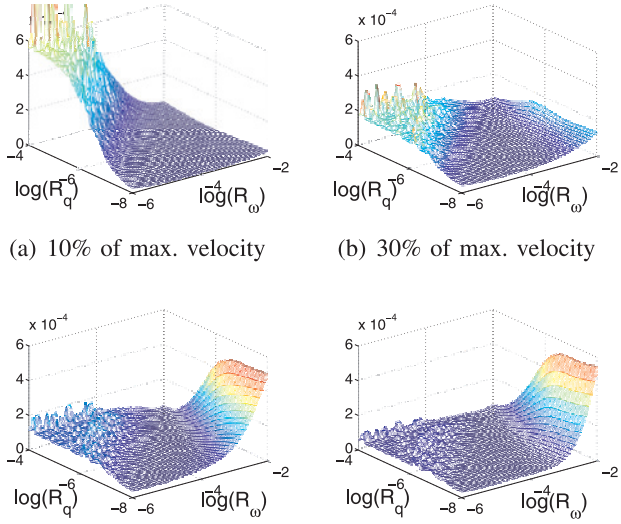
Figures 6 and 7 show the performance indexes for different covariance values of $\mathbf{R_q}$ and $\mathbf{R}_\omega$ related to the rotational movements using EKF and UKF estimation. It can be appreciated that for low speed motions the estimation decreases (index increases) with assumed low uncertainty on $\boldsymbol{\omega}$ and high uncertainty on $\mathbf{q}$. Thus, it is shown that, at low speeds, the estimation relies on the vision system, as expected. In contrast, for high speed motions, estimation mainly relies on inertial data. It is interesting to remark that the numerical results do not just demonstrate that vision is crucial for low motion movements and inertial is very important for faster movements, but also give guidelines regarding the tuning process of covariance values in terms of accuracy of the estimated motions. These results highlight the benefits of fusing both inertial and vision data, since the sensors are complementary, and good performance can be obtained when an intermediate trade-off is considered for both sensors. Outside of the calculated bounds of covariance, the estimation becomes numerically unstable in most cases, and they are not reported. It can also be appreciated that the UKF presents some peaks in the performance index, which do not appear in the EKF, which indicates that UKF has worse numeric behavior than EKF.



(a) 10% of max. velocity (b) 30% of max. velocity

(c) 50% of max. velocity (d) 75% of max. velocity

Fig. 6. Covariance influence over the performance index J_q under different rotational movements with EKF.

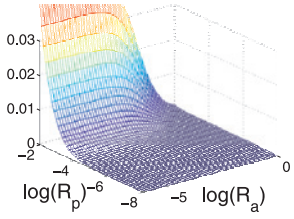


(a) 10% of max. velocity (b) 30% of max. velocity

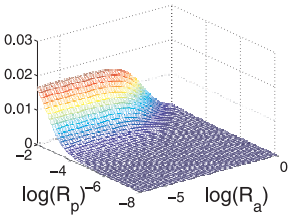
(c) 50% of max. velocity (d) 75% of max. velocity

Fig. 7. Covariance influence over the performance index J_q under different rotational movements with UKF.

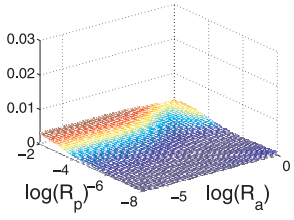
Similarly, Figures 8 and 9 show the performance index for different covariance values of $\mathbf{p}$ and $\mathbf{a}$, related to translational movements. For low speed experiments (10% and 30%) there is a region with higher covariance values for vision and lower for inertial where the estimation is very poor. This is because the influence of vision is lower and consequently biases of accelerations are not correctly estimated. Slow motions are affected by biases of acceleration measurements, while for fast motions, they can be neglected. The main difference with re-



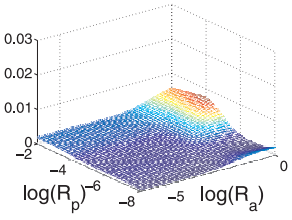
(a) 10% of max. velocity



(b) 30% of max. velocity

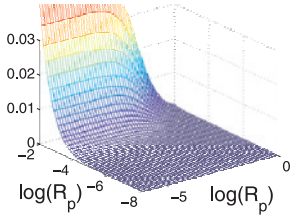


(c) 50% of max. velocity

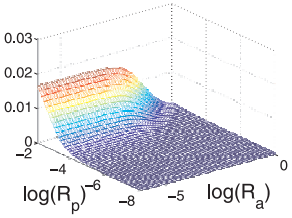


(d) 75% of max. velocity

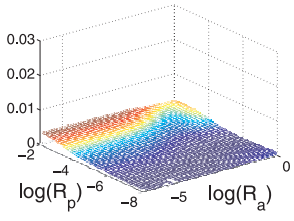
Fig. 8. Covariance influence on the performance index J_p under different translational movements with EKF.



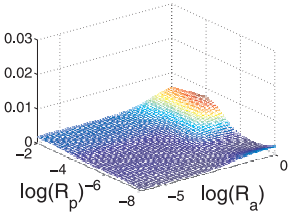
(a) 10% of max. velocity



(b) 30% of max. velocity



(c) 50% of max. velocity

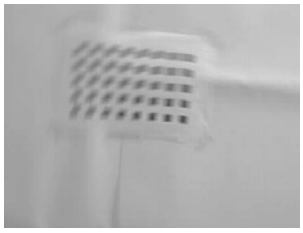


(d) 75% of max. velocity

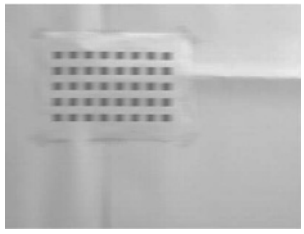
Fig. 9. Covariance influence on the performance index J_p under different translational movements with UKF.

spect to rotational movements is that, in this case, vision is much more crucial, since accelerations require a double integration to compute the pose, while angular velocities require only a single integration to compute the orientation.

Based on the results obtained we have selected the following values for $\mathbf{R}_k$ and $\mathbf{Q}_k$ common to all velocities:



(a) Rotational Mov. at 75%.



(b) Translational Mov. at 75%.

Fig. 10. Blurred images examples.

$$\mathbf{R}_k = \text{diag}\left\{ \begin{array}{l} 10^{-3}\cdot\mathbf{I}_{3\times3}, 10^{-4}\cdot\mathbf{I}_{3\times3}, \\ 10^{-7}\cdot\mathbf{I}_{3\times3}, 10^{-6}\cdot\mathbf{I}_{4\times4} \end{array} \right\} \quad (25)$$

$$\mathbf{Q}_k = \text{diag}\left\{ \begin{array}{l} 0.7447\cdot\mathbf{I}_{3\times3}, 0.38\cdot\mathbf{I}_{3\times3}, \\ 0.19\cdot10^{-6}\cdot\mathbf{I}_{3\times3} \end{array} \right\}. \quad (26)$$

Figures 11(a), 11(b), 11(c) and 11(d) show the estimation results for rotational movements using EKF and UKF. Both filters give nearly the same results for selected covariance values, and therefore their responses overlap.

In Figures 11(a) to 11(d), vision measurements are represented with dots; large gaps between dots are due to failures of the detection process of the vision system, which are much more frequent with faster motions. Figure 10 shows the captured blurred image from the vision system on a detection failure. It can be appreciated that the number of detected features is too low to provide an accurate pose estimation and therefore the image is rejected. Despite this fact, the fusion integration of inertial measurements aims to re-construct the correct motion.

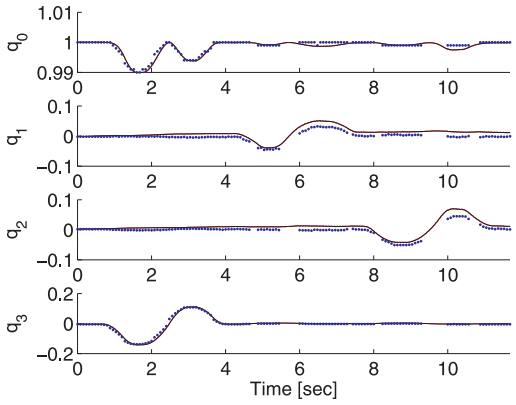
Similarly, results for translational movements are depicted in Figures 12(a), 12(b), 12(c) and 12(d), where the same conclusions as before can be derived.

According to these results, for this particular application, EKF gives better performance than UKF, since both filters provide nearly the same estimation, but the computational cost of UKF is about 7 times higher.

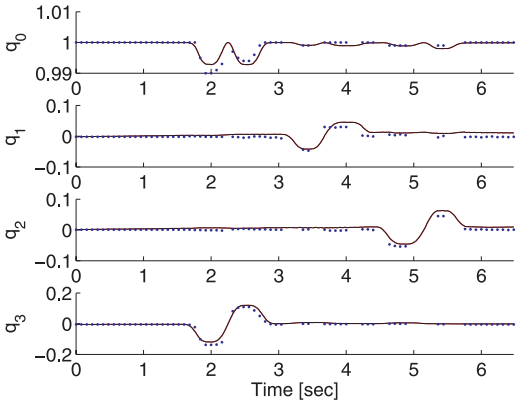
Finally, a study of the benefits of fusing vision and inertial measurements is performed. In that sense, the performance index is also calculated for pure inertial or pure vision. Figure 13 shows that the fusion of inertial and vision gives better performance results than single estimations. Fusion introduces more benefits to pure inertial than to pure vision estimation. This is mainly due to the double integration performed on the inertial measurements where bias correction could not be performed.

5. Conclusions

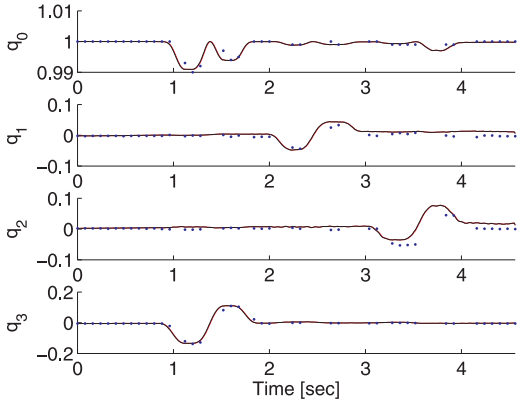
The tracking of fast movements is a difficult task, particularly if it is performed with a pure vision or pure inertial system.



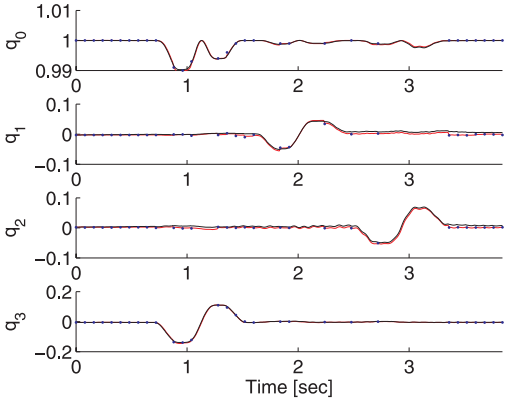
(a) Rotational movement est. at 10% of max. velocity.



(b) Rotational movement est. at 30% of max. velocity.



(c) Rotational movement est. at 50% of max. velocity.



(d) Rotational movement est. at 75% of max. velocity.

Fig. 11. Estimation results of rotational movements (continuous lines) and with vision measurements (dots).

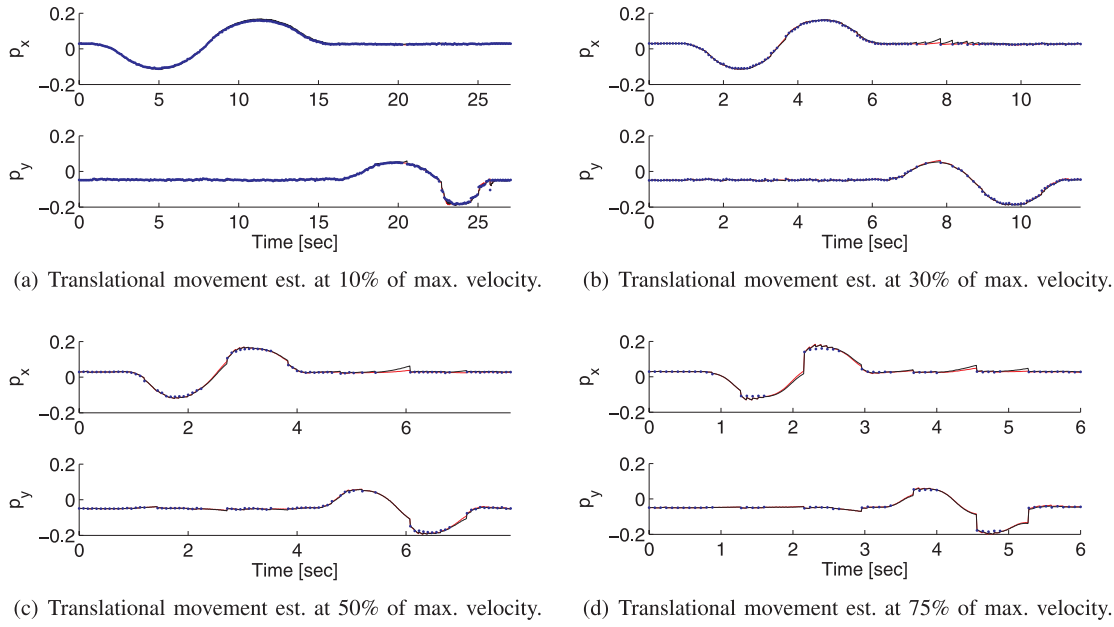
This problem can be solved if both sensors are fused, since they provide complementary properties.

In this paper, we have presented a tracking system for ego-motion estimation, which can recover vision failures during fast motions. The fusion is performed by considering a EKF and UKF with multi-rate sampling of measurements. They have shown to be robust enough for this application and therefore they have been tested alongside other non-Gaussian filters. In both fusion techniques (EKF and UKF), each sensor is sampled at the highest frequency at which they can provide measurements. The approach considered in this paper uses multi-rate holds and samplers to interface signals at different frequencies. Estimations with MR-EKF (multi-rate EKF) and MR-UKF (multi-rate UKF) provided very similar results, without significant differences between them. The computational cost of UKF is about 7 times higher than the cost of EKF computation.

This approach has not only been validated in Vision/Inertial fusion, but also has been recently validated in Laser/Encoder fusion for mobile robot self-localization and map building using an asynchronous multi-rate FastSLAM (Armesto et al 2007).

In addition to this, the paper investigates the influence of the covariance matrices of noise. Conclusions from this analysis lead to determination of appropriate tuning values between vision and inertial measurements. This aspect is crucial since it has a direct influence on the estimation performance. It has been shown that a common set of covariance values exist that give good performance over a range of motion speeds. A set-up based on an industrial robot arm has been used to validate the estimation. This set-up allows us to pre-define basic rotational and translational motions, which can be combined to generate complex motions.

Future research is oriented towards estimating even more complex motions such as natural human movements. In addition to this, other fusion techniques such as Particle Filters will be tested as well as SLAM techniques to estimate the pose and the structure (map). In the context of the RESTAURO Re-



(a) Translational movement est. at 10% of max. velocity.

(b) Translational movement est. at 30% of max. velocity.

(c) Translational movement est. at 50% of max. velocity.

(d) Translational movement est. at 75% of max. velocity.

Fig. 12. Estimation results for translational movements (continuous lines) and with vision measurements (dots).

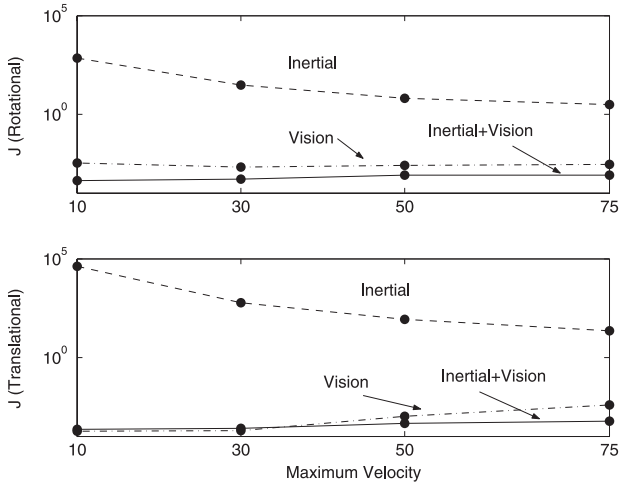


Fig. 13. Mean quadratic error versus movement power.

search Project, fusion of Vision/GPS/Inertial data will be combined for better estimation, improving such required tasks such as collision detection and avoidance, defects detection and recovery, and feedback position control.

Appendix

The multimedia extensions to this article can be found online by following the hyperlinks from www.ijrr.org.

Table 2. Multimedia Extensions of Fast Ego-Motion Estimation with Multi-rate Fusion of Inertial and Vision.

Extension	Media Type	Description
Extension 1	Data	Raw-data of experiments.
Extension 2	Code	Matlab code.
Extension 3	Videos	Videos of pre-defined movements

References

Abuhadrous, I., Nashashibi, F., and Laurgeau, C. (2003). 3D land vehicle localization: a real-time multi-sensor data fusion approach using RT MAPS. *International Conference on Advanced Robotics*, pp. 71–76.

Albertos, P. (1990). Block multirate input-output model for sampled-data control systems. *IEEE Transactions on Automatic Control*, **AC-35**(9): 1085–1088.

Alves, J., Lobo, J., and Dias, J. (2003). Camera-inertial sensor modelling and alignment for visual navigation. *International Conference on Advanced Robotics*, pp. 1693–1698.

Armesto, L. I. (2007). An asynchronous multi-rate approach to probabilistic self-localisation and mapping. Accepted in *Robotics and Automation Magazine*.

Armesto, L. and Tornero, J. (2003). Dual-rate high order holds based on primitive functions. *American Control Conference*, pp. 1140–1145.

Armesto, L. and Tornero, J. (2004). SLAM based on Kalman filter for multirate fusion of laser and encoder measurements. *IEEE International Conference on Intelligent Robots and Systems*, pp. 1860–1865.

Armesto, L., Chroust, S., Vincze, M., and Tornero, J., (2004). Multi-rate fusion with vision and inertial sensors. *International Conference on Robotics and Automation*, pp. 193–199.

Bradski, G. (2000). An open-source library for processing image data. *Dr Dobbs Journal*, November.

Carpenter, J., Clifford, P., and Fernhead, P. (1997). An improved particle filter for non-linear problems. Technical Report, Department of Mathematics, Imperial College.

Chai, L., H., Hoff, W.A., and Vincent, T. (2002). Three-dimensional motion and structure estimation using inertial sensors and computer vision for augmented reality. *Presence: Teleoperators and Virtual Environments*, **11**(5): 474–492.

Chou, J. (1992). Quaternion kinematic and dynamic differential equations. *IEEE Transactions on Robotics and Automation*, **8**(1): 53–64.

Chroust, S. and Vincze, M. (2003). Fusion of vision and inertia data for motion and structure estimation. *Journal of Robotics Systems*, **21**(2): 73–83.

Dissanayake, M, Newman, P., Clark, S., Durrant-Whyte, H., and Csorba, M. (2001). A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotics and Automation*, **17**(3): 229–241.

Doucet, A., de Freitas, N., Murphy, K., and Russell, S. (2000). RaoBlackwellised particle filtering for dynamic Bayesian networks. In Uncertainty in Artificial Intelligence.

Doucet, A., Gordon, N., and Krishnamurthy, V. (2001). Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, **49**(3): 613–624.

Gemeiner, P., Armesto, L., Montes, N., Tornero, J., Vincze, M., and Pinz, A. (2006). Visual tracking can recapture after fast motions with the use of inertial sensors. *Digital Imaging and Pattern Recognition, OAGM/AAPR*, pp. 141–150.

Goddard, J. and Abidi, M. (1998). Pose and motion estimation using dual quaternion-based extended Kalman filtering. *SPIE*, **3313**: 189–200.

Goodwin, G. and Feuer, A. (1992), Linear periodic control: A frequency domain viewpoint. *Systems and Control Letters*, **19**: 379–390.

Gordon, N., Salmond, D., and Smith, A. (1993). Novel aproach to nonlinear/non-gaussian Bayesian state estimation. *IEE-Proceedings-F*, **140**(2): 107–113.

Grewal, M. S., Weill, L. R., and Andrews, A. P. (2001). *Global Positioning Systems, Inertial Navigation, and Integration*. John Wiley and Sons, Canada.

Gurfil, P and Kasdin, N. (2002). Two-step optimal estimator for three dimensional target tracking. *American Control Conference*, pp. 209–214.

Huster, A. and Rock, S. (2003). Relative position sensing by fusing monocular vision and inertial rate sensors. *International Conference on Advanced Robotics*, pp. 1562–1567.

Huster, A. and Rock, S. (2001). Relative position estimation for interventioncapable AUVS by fusing vision and inertial measurements. *International Symposium on Unmanned Unthethered Submersible Technology*.

Jekeli, C. (2001). *Inertial Navigation Systems with Geodetic Applications*. Walter de Gruyter. John Wiley & Sons, Canada.

Julier, S. and Uhlmann, J. (2002), Reduced sigma points filters for the propagation of means and covariances through nonlinear transformations. *American Control Conference*, Vol. 2, pp. 887–892.

Khargonekar, P., Poolla, K., and Tannenbaum, A. (1985). Robust control of linear time-invariant plants using periodic compensation. *IEEE Transactions on Automatic Control*, **AC-30**: 1088–1985.

Kranc, G. (1957). Input-output analysis of multirate feedback systems. *IEEE Transactions on Automatic Control*, **AC-3**: 21–28.

Lobo, J. and Dias, J. (1998). Integration of inertial information with vision. *IEEE Industrial Electronic Society*, pp. 1263–1267.

Longhi, S. (1994). Structural properties of multirate sampled systems. *IEEE Transactions on Automatic Control*, **39**(3): 692–696.

Panerai, F., Metta, G., and Sandini, G. (2000). Visuo-inertial stabilization in space-variant binocular systems. *Robotics and Autonomouns Systems*, **30**(1-2): 195–214.

Rehbinder, H. and Ghosh, B. (2001). Multi-rate fusion of visual and inertial data. *International Conference on Multi-Sensor Fusion and Integration for Intelligent Systems*, pp. 97–102.

Smith, A. and Gelfand, E. (1992). Bayesian statistics without tears: A samplingresampling perspective. *American Statistician*, **2**: 84–88.

Tornero, J. (1985). Non-conventional sampled-data systems modelling. University of Manchester (UMIST), Control System Center Report, 640/1985.

Tornero, J., Albertos, P., and Salt J. (1999). Periodic optimal control of multirate sampled data systems. *14th World Congress of IFAC*, China, pp. 211–216.

Tornero, J., Gu, Y., and Tomizuka, M. (1999). Analysis of multi-rate discrete equivalent of continuous controller. *American Control Conference*, pp. 2759–2763.

Tornero, J. and Tomizuka, M. (2000). Dual-rate high order hold equivalent controllers. *American Control Conference*, pp. 175–179.

Tornero, J. and Tomizuka, M. (2002). Modeling, analysis and design tools for dual-rate systems. *American Control Conference*, pp. 4116–4121.

Ude, A. (1999). Filtering in a unit quaternion space for model-based object tracking. *IEEE Transactions Robotics and Autonomous Systems*, **28**(2–3): 163–172.

Vincze, M., Ayromlou, M., Ponweiser, W., and Zillich, M. (2001). Edge projected integration of image and model cues for robust model-based object tracking. *International Journal of Robotics Research*, **7**(20): 533–552.

Zhang, Z. (1998). A flexible new technique for camera calibration. Technical Report, Microsoft Research, Microsoft Corporation.