2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Visual and Inertial Multi-Rate Data Fusion for Motion Estimation via Pareto-Optimization

Giuseppe Loianno[1], Vincenzo Lippiello[1], Carlo Fischione[2], and Bruno Siciliano[1]

Abstract— **Motion estimation is an open research field in control and robotic applications. Sensor fusion algorithms are generally used to achieve an accurate estimation of the vehicle motion by combining heterogeneous sensors measurements with different statistical characteristics. In this paper, a new method that combines measurements provided by an inertial sensor and a vision system is presented. Compared to classical model-based techniques, the method relies on a Pareto optimization that trades off the statistical properties of the measurements. The proposed technique is evaluated with simulations in terms of computational requirements and estimation accuracy with respect to a classical Kalman filter approach. It is shown that the proposed method gives an improved estimation accuracy at the cost of a slightly increased computational complexity.**

I. Introduction

Localization is an essential feature in robotic applications, such as surveillance and monitoring. When a robot moves in unknown environments information about its current position is needed. The use of filters combining different sensor data, which are generally provided at different sampling rates, is highly appealing to make an accurate localization.

Different methods have been proposed to combine heterogeneous data sources such as Global Positioning System (GPS), Inertial Measurement Unit (IMU), odometry sensors, and local radio technologies [1], [2]. Nevertheless, this remains an open research field in robotics and especially in Micro Aerial Vehicles (MAVs) applications, where low cost IMUs have to be combined with camera(s) information, as well as with GPS data. However, due to the integration error, position and velocity can be estimated for no more than few seconds by using only IMU data. On the other hand, vision systems provide positional information with no drift with respect to fixed observed environments. The main drawback of this sensors is the huge amount of data to be elaborated, that generates time delay in the estimation update and low measurement rate compared to IMU sensors.

Unequal sampling times of the measurement devices rises significant challenges. A solution consists in adopting multi-rate filters [3]. Further, the delay that characterizes visual measurements can be addressed by adopting techniques based on Kalman filters or its variants [4], [5]. However, the delay compensation is often achieved by a state augmentation depending on the given delay [6]. In [7] the general Kalman filter formulation is extended by considering both the relative measurements update and the correlation between two consecutive displacements, while a solution to choose the initial state covariance matrix is addressed in [8]. The solution proposed in [3] has been employed in [9] by compensating the delay due to the wireless data communication and image processing to stabilize a MAV with a standard PID controller.

The use of mono-camera systems in an unknown environment allows the direct estimation of the vehicle egomotion up to a scale factor. By combining inertial and visual data, the global scale factor can be estimated achieving an absolute egomotion estimation. The solutions proposed in [10], [11], [12] combine inertial measurements and consecutive feature matchings obtaining a full-scale closed-form solution.

Optimal sensor-fusion techniques based on second-order moment minimization [13], [14] and Pareto Optimization [15] attempt to couple heterogeneous sensors such as Ultra-Wideband radio measurements with speed and absolute orientation information. Other works rely on the use of complementary filters and non-linear estimators as in [16], [17]. In these latter cases, the vehicle position, velocity and attitude estimation is obtained with a non-linear dynamic system.

In this paper a new optimal sensor-fusion algorithm based on Pareto optimization techniques is proposed to combine IMU and camera visual measurements to estimate a vehicle motion. The advantages of the proposed method are the no-prior assumption about robot motion model and that the multi-rate sensor fusion. A comparison of the proposed technique with a Kalman filter approach shows an improved estimation at the price of a limited increased computational complexity.

II. Problem Formulation

The position estimation of a flying vehicle equipped with an IMU and a differential visual system is the main goal of this paper. Without loss of generality, the world reference frame is assumed coincident with the first IMU pose, and the orientation of the camera with respect to the IMU is supposed to be known. The differential vision system provides the position displacement with respect to the previous vehicle position, i.e. differential positional measurement is available, with a sampling rate T_V. On the other hand, the

This work was supported by the ARCAS and SHERPA collaborative projects, which have received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreements ICT-287617 and ICT-600958, respectively. The authors are solely responsible for its content. It does not represent the opinion of the European Community and the Community is not responsible for any use that might be made of the information contained therein.

[1]The authors are with PRISMA Lab, Department of Electrical Engineering and Information Technology, University of Naples Federico II, via Claudio 21, 80125, Naples, Italy {giuseppe.loianno, vincenzo.lippiello, bruno.siciliano}@unina.it

[2]The author is with ACCESS Linnaeus Center, School of Electrical Engineering, KTH, Royal Institute of Technology, Stockholm, Sweden carlofi@kth.se

IMU provides the linear acceleration and attitude with a sampling time $T_i \leq T_V$. The latter can be recovered by fusing the accelerometer and rate gyro measurements in a standard complementary attitude filter [18], by using a gradient descent method [19]. The orientation of the vehicle is represented with the Tait-Bryan angles of roll, pitch, and yaw $\boldsymbol{\phi} = (\varphi, \theta, \psi)$, which yields to the rotation matrix $\boldsymbol{R}(\boldsymbol{\phi}) \in SO(3)$, i.e.

$$\boldsymbol{R}(\boldsymbol{\phi}) = \begin{bmatrix} c_\varphi c_\theta & c_\varphi s_\theta s_\psi - s_\varphi c_\psi & s_\varphi s_\psi + c_\varphi s_\theta c_\psi \\ s_\varphi c_\theta & s_\varphi s_\theta s_\psi + c_\varphi c_\psi & s_\varphi s_\theta c_\psi - c_\varphi s_\psi \\ -s_\theta & c_\theta s_\psi & c_\theta c_\psi \end{bmatrix},$$

where $s_x = \sin(x)$ and $c_x = \cos(x)$. Moreover, the measurement of the Euler angles

$$\tilde{\boldsymbol{\phi}} = \left(\tilde{\varphi}, \tilde{\theta}, \tilde{\psi}\right) \tag{1}$$

can be modeled as

$$\begin{aligned} \tilde{\varphi} &= \varphi + \omega_\varphi \\ \tilde{\theta} &= \theta + \omega_\theta \\ \tilde{\psi} &= \psi + \omega_\psi, \end{aligned} \tag{2}$$

where ω_φ, ω_θ, and ω_ψ represent angle Gaussian white noises with zero mean and variance σ_ϕ^2, σ_θ^2, and σ_ψ^2, respectively.

Finally, the linear acceleration of the vehicle $\tilde{\boldsymbol{a}}$ with respect to the fixed frame is given by

$$\tilde{\boldsymbol{a}} = \boldsymbol{R}(\tilde{\boldsymbol{\phi}})\tilde{\boldsymbol{a}}_I = \boldsymbol{R}(\tilde{\boldsymbol{\phi}})\left(\boldsymbol{a}_I + \boldsymbol{\omega}_{a_i}\right) - \boldsymbol{g}, \tag{3}$$

where $\boldsymbol{a}_I = [\ a_{I,x} \quad a_{I,y} \quad a_{I,z}\ ]^{\mathrm{T}}$ is the acceleration provided by the IMU and expressed in the current robot frame, $\boldsymbol{\omega}_{a_I} = [\ \omega_{a_I,x} \quad \omega_{a_I,y} \quad \omega_{a_I,z}\ ]^{\mathrm{T}}$ is a Gaussian white noise with known variance $\boldsymbol{\sigma}^2_{a_I} = [\ \sigma^2_{a_I,x} \quad \sigma^2_{a_I,y} \quad \sigma^2_{a_I,z}\ ]^{\mathrm{T}}$ and $\boldsymbol{g} = [\ 0 \quad 0 \quad 9.81\ ]^{\mathrm{T}}$ is the gravity vector that can be subtracted from the inertial measurement given that the absolute orientation is known.

The visual measurement $\widetilde{\delta \boldsymbol{p}_V} = \delta \boldsymbol{p} + \boldsymbol{\omega}_V$ represents the robot displacement performed during the last sampling period T_V, expressed in the current frame, where $\delta \boldsymbol{p}$ is the effective unknown displacement and $\boldsymbol{\omega}_V = [\ \omega_{V,x} \quad \omega_{V,y} \quad \omega_{V,z}\ ]^{\mathrm{T}}$ is white noise with known variance $\boldsymbol{\sigma}^2_V = [\ \sigma^2_{V,x} \quad \sigma^2_{V,y} \quad \sigma^2_{V,z}\ ]^{\mathrm{T}}$ and bias $\boldsymbol{b}_V = [\ b_{V_x} \quad b_{V_y} \quad b_{V_z}\ ]^{\mathrm{T}}$. Known visual odometry techniques (e.g. see [20] and [21]) can be considered to compute $\widetilde{\delta \boldsymbol{p}_V}$.

With the proposed approach, the estimation of the vehicle displacement $\widehat{\delta \boldsymbol{p}} = [\ \widehat{\delta x} \quad \widehat{\delta y} \quad \widehat{\delta z}\ ]^{\mathrm{T}}$, between the sampling times $k-1$ and k, is evaluated by a convex combination of the visual measurement and the inertial displacement estimation:

$$\widehat{\delta x}(k) = (1 - \beta_{x,k})\widetilde{\delta x}_V(k) + \beta_{x,k}\widehat{\delta x}_a(k) \tag{4}$$

$$\widehat{\delta y}(k) = (1 - \beta_{y,k})\widetilde{\delta y}_V(k) + \beta_{y,k}\widehat{\delta y}_a(k) \tag{5}$$

$$\widehat{\delta z}(k) = (1 - \beta_{z,k})\widetilde{\delta z}_V(k) + \beta_{z,k}\widehat{\delta z}_a(k). \tag{6}$$

The term $\widehat{\delta \boldsymbol{p}}_a = [\ \widehat{\delta x}_a \quad \widehat{\delta y}_a \quad \widehat{\delta z}_a\ ]^{\mathrm{T}}$ is the estimation of the position displacement obtained from the inertial data.The weight factors $\beta_{x,k}$, $\beta_{y,k}$, and $\beta_{z,k}$ are the unknown parameters that have to be (optimally) chosen at each sampling time by a Pareto optimization, as described in the following.

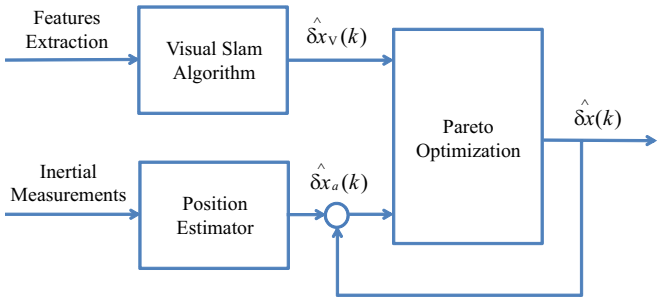


Fig. 1. Working schema of the proposed Pareto optimization algorithm

Accordingly, the absolute position estimation of the vehicle at the time instant k can be computed as

$$\widehat{\boldsymbol{p}}(k) = \widehat{\boldsymbol{p}}(k - N) + \widehat{\delta \boldsymbol{p}}(k), \tag{7}$$

where $N = 1$ in case of synchronous measurements.

Without loss of generality, only the estimation of the x component will be described. Analogous results can be achieved for the y and z motion components.

Figure 1 shows the working principle of the sensor fusion algorithm valid both in synchronous and asynchronous measurements cases. The difference between the two cases relies on the correction on inertial position estimation that, as it will shown in next subsections, in the synchronous case is done at every step time, while in the asynchronous case only where there is visual measurement availability.

A. Synchronous measurements

Consider the case of synchronous measurements, $T = T_V = T_i$. Then the IMU measurements are used at the same frequency of the vision system. By starting from the inertial measurements, we obtain the estimation of the positional displacement by a forward Euler integration

$$\widehat{v}_x(k) = \widehat{v}_x(k-1) + T\tilde{a}_x(k) \tag{8}$$

$$\widehat{\delta x}_a(k) = T\widehat{v}_x(k). \tag{9}$$

By plugging (8) into (9) and approximating $\widehat{v}_x(k-1) \cong \widehat{\delta x}(k-1)/T$, the estimation of the position displacement obtained by the inertial measurement is

$$\widehat{\delta x}_a(k) = \widehat{\delta x}(k-1) + T^2\tilde{a}_x(k). \tag{10}$$

Thus, the estimated position depends on the estimated position at time $k-1$ and on the acceleration measurement.

B. Asynchronous measurements

In the asynchronous case, the vision system provides the vehicle pose estimation at lower frequency with respect to the inertial system, $T_V = NT_i$, with $N \in \mathbb{N}$ being the scale factor relating the IMU and vision frequencies. By exploiting classical Taylor expansion the position at time k can be written as a function of the acceleration a_x at time $k-1$:

$$x(k) = x(k-1) + T_i\delta v_x(k-1) + \frac{T_i^2}{2}a_x(k-1), \tag{11}$$

where δv_x represents the velocity at instant time $k-1$.

Thus, the absolute position estimation obtained by the inertial sensor at time instant k can be expressed as a function of the optimal position estimation at time $k-N$ as

$$\widehat{x}_a(k) = \widehat{x}(k-N) + T_i \sum_{j=k-N}^{k-1} \widehat{\delta v}_x(j) + \frac{T_i^2}{2} \sum_{j=k-N}^{k-1} a_x(j), \tag{12}$$

where the velocity estimation is achieved with the Euler integration $\widehat{\delta v}_x(j) = \widehat{\delta v}_x(j-1) + T_i \tilde{a}_x(j-1)$. Hence, the differential displacement from inertial measurement between two consecutive optimization time instants is obtained by subtracting $\widehat{x}_a(k-N)$ to (12):

$$\begin{aligned}\widehat{\delta x}_a(k) =& \widehat{\delta x}(k-N) + T_i \sum_{j=k-N}^{k-1} \widehat{\delta v}_x(j) + \frac{T_i^2}{2} \sum_{j=k-N}^{k-1} a_x(j) \\ &- T_i \sum_{j=k-2N}^{k-N-1} \widehat{\delta v}_x(j) - \frac{T_i^2}{2} \sum_{j=k-2N}^{k-N-1} a_x(j),\end{aligned} \tag{13}$$

The state estimation $\widehat{\delta x}(k)$ will be performed as soon the visual measurement is available, according to (4).

The differential position estimation provided by the vision system is assumed coincident with the measurement itself

$$\widetilde{\delta x}_V(k) = \widehat{\delta x}_V(k) = \delta x_V(k) + \omega_{x_V}(k), \tag{14}$$

where $\delta x_V(k)$ is the ground-truth position displacement.

C. Pareto optimization problem

The estimator (4) can be modeled as follows

$$\widehat{\delta x}(k) = \delta x(k) + \omega_x(k), \tag{15}$$

where $\omega_x(k)$ is the error position, which is estimated as described in equation (3). The estimator bias is denoted by

$$P_1 = \mathbb{E}\{\omega_x(k)\}, \tag{16}$$

and the estimation variance is defined by

$$P_2 = \mathbb{E}\{(\widehat{\delta x}(k) - \mathbb{E}\{\widehat{\delta x}(k)\})^2\}, \tag{17}$$

where $\mathbb{E}\{\cdot\}$ is the expected value of a random variable.

Then we can pose a Pareto optimization problem as

$$\begin{aligned}&\min_{\beta_{x,k}} (1-\rho_{x,k})\, P_2 + \rho_{x,k} P_1^2 \\ &\text{s.t. } \beta_{x,k} \in (-1,1);\end{aligned} \tag{18}$$

with this choice, the bias and the variance of the estimation error will be minimized simultaneously. The Pareto weighting factor $\rho_{x,k}$ has to be chosen at each step so to trade-off the high variance and bias of sensors. The constraint on $\beta_{x,k}$ is required because the bias may become unstable with time when the statistical modeling of the bias is computed, as we show later. The solution of the optimization problem requires first the evaluation of the quantities P_1 and P_2, that will be discussed in the following sections.

III. Error Estimation Bias and Variance

In this section an analytical recursive expression for Eqs. (16) and (17) is provided. The computation of the bias and of the variance will be differentiated according to the types of signals to be fused and the sensor timing condition, i.e. synchronous/asynchronous cases.

A. Synchronous measurements

By considering (4) and (15), the error $\omega_x(k)$ becomes

$$\begin{aligned}\omega_x(k) =& (1-\beta_{x,k})\omega_V(k) + \beta_{x,k}\left(\widehat{\delta x}_a(k) - \delta x_a(k)\right) \\ =& (1-\beta_{x,k})\omega_V(k) + \beta_{x,k}(\omega_x(k-1) \\ &+ T^2 \omega_{a_x}(k-1)).\end{aligned} \tag{19}$$

Consequently, (16) can be rewritten as

$$\begin{aligned}P_1 =& \mathbb{E}\{\omega_x(k)\} = (1-\beta_{x,k})\mathbb{E}\{\omega_{V_x}(k)\} \\ &+ \beta_{x,k}\mathbb{E}\{\omega_x(k-1))\} + \beta_{x,k} T^2 \mathbb{E}\{\omega_{a_x}(k-1)\},\end{aligned} \tag{20}$$

where

$$\begin{aligned}&\mathbb{E}\{\omega_{a_x}(k)\} = a_{I,x}(k) c_\varphi c_\theta e^{\frac{-\sigma_\varphi^2}{2}} e^{\frac{-\sigma_\theta^2}{2}} + a_{I,y}(k) \\ &(c_\varphi s_\theta s_\psi e^{\frac{-\sigma_\varphi^2}{2}} e^{\frac{-\sigma_\theta^2}{2}} e^{\frac{-\sigma_\psi^2}{2}} - s_\varphi c_\psi e^{\frac{-\sigma_\varphi^2}{2}} e^{\frac{-\sigma_\psi^2}{2}}) + a_{I,z}(k) \\ &(s_\varphi s_\psi e^{\frac{-\sigma_\varphi^2}{2}} e^{\frac{-\sigma_\psi^2}{2}} + c_\varphi s_\theta c_\psi e^{\frac{-\sigma_\varphi^2}{2}} e^{\frac{-\sigma_\theta^2}{2}} e^{\frac{-\sigma_\psi^2}{2}}) - a_{I,x}(k) c_\varphi c_\theta \\ &- a_{I,y}(k)(c_\varphi s_\theta s_\psi - s_\varphi c_\psi) - a_{I,z}(k)(s_\varphi s_\psi + c_\varphi s_\theta c_\psi).\end{aligned}$$

The term $\mathbb{E}\{\omega_{a_x}(k)\}$ is derived as in [13], [15]

$$\begin{aligned}\mathbb{E}\{\widetilde{a}(k)\cos(\widetilde{\varphi}(k))\} &= a(k)\cos(\varphi(k)) e^{\frac{-\sigma_\varphi^2}{2}} \\ \mathbb{E}\{\widetilde{a}(k)\sin(\widetilde{\varphi}(k))\} &= a(k)\sin(\varphi(k)) e^{\frac{-\sigma_\varphi^2}{2}},\end{aligned}$$

where $\tilde{\boldsymbol{\phi}}(k)$ has been considered Gaussian and statistically independent in its own components and with respect to the acceleration measurements $\boldsymbol{a}_I$ (see Appendix A).

In the proposed formulation the bias on the single acceleration measurement has been neglected for simplicity. In fact, since it is generally constant, it can be estimated and subtracted from the measurement itself [10]. An alternative approach can be to keep the bias in the measurement and consider it as a penalty in P_1 for the acceleration pose estimation, thus obtaining a formulation similar to the one presented above.

Notice that (20) is a discrete-time recursive expression, where the value of velocity bias at time k is related to that one at time $k-1$ trough $\beta_{x,k}$ coefficient. This can be directly interpreted as a discrete time differential equation. To avoid a blow up of the bias, all the eigenvalues should be in the circle of radius 1. Hence, it is necessary that $|\beta_{x,k}| \in (0,1)$ and thus $\beta_{x,k} \in (-1,1)$, as required in (18).

By substituting (20) in (17), P_2 can be rewritten as

$$\begin{aligned}P_2 =& \mathbb{E}\{(\widehat{\delta x}(k) - \mathbb{E}\{\widehat{\delta x}(k)\})^2\} = \mathbb{E}\{((1-\beta_{x,k})v_{V_x}(k) \\ &+ \beta_{x,k} v_x(k-1) + \beta_{x,k} v_{a_x}(k-1))^2\},\end{aligned} \tag{21}$$

where

$$\begin{aligned} v_{V_x}(k) &\triangleq \omega_{V_x}(k) - \mathbb{E}\{\omega_{V_x}(k)\} \\ v_x(k) &\triangleq \omega_x(k) - \mathbb{E}\{\omega_x(k)\} \\ v_{a_x}(k) &\triangleq \omega_{a_x}(k) - \mathbb{E}\{\omega_{a_x}(k)\} \\ v_{v_x}(k) &\triangleq \omega_{v_x}(k) - \mathbb{E}\{\omega_{v_x}(k)\}, \end{aligned} \tag{22}$$

with $\sigma^2_{V_x}$, σ^2_x, $\sigma^2_{a_x}$ and $\sigma^2_{v_x}$ the corresponding variances. Since v_{V_x}, v_x, v_{a_x} and v_{v_x} are statistically independent and $\mathbb{E}\{v_{a_x}\} = \mathbb{E}\{v_x\} = 0$, (21) yields

$$\begin{aligned} P_2 =& (1-\beta_{x,k})^2\sigma^2_{V_x}(k) + \beta^2_{x,k}\sigma^2_x(k-1) \\ &+ \beta^2_{x,k}T^4\sigma^2_{a_x}(k-1), \end{aligned} \tag{23}$$

where

$$\begin{aligned} \sigma^2_{a_x}(k) =& \mathbb{E}\{(\omega_{a_x}(k) - \mathbb{E}\{\omega_{a_x}(k)\})^2\} = \mathbb{E}\{\omega^2_{a_x}(k) + \\ & \mathbb{E}^2\{\omega_{a_x}(k)\} - 2\omega_{a_x}(k)\,\mathbb{E}\{\omega_{a_x}(k)\}\} \\ =& \mathbb{E}\{\omega^2_{a_x}(k)\} - \mathbb{E}^2\{\omega_{a_x}\}. \end{aligned}$$

The second order moment $\mathbb{E}\{\omega^2_{a_x}(k)\}$ could be characterized using the following properties [13], [15]:

$$\begin{aligned} \mathbb{E}\{\widetilde{a}^2(k)\cos^2(\widetilde{\varphi}(k))\} &= \sigma^2_a\left(\frac{1}{2} + \frac{1}{2}c_{2\varphi}e^{-2\sigma^2_\varphi}\right) \\ \mathbb{E}\{\widetilde{a}^2(k)\sin^2(\widetilde{\varphi}(k))\} &= \sigma^2_a\left(\frac{1}{2} - \frac{1}{2}c_{2\varphi}e^{-2\sigma^2_\varphi}\right), \end{aligned}$$

where σ^2_a is the acceleration variance (see Section A).

B. Asynchronous measurements

By considering equation (4), (13), and (15) the quantity $\omega_x(k)$ can be expressed as follows

$$\omega_x(k) = (1-\beta_{x,k})\omega_V(k) + \beta_{x,k}\omega_{x_a}(k), \tag{24}$$

where $\omega_{x_a}(k)$ represents the position differential error of the inertial measurements in (13). Hence, (16) can be written as

$$\begin{aligned} P_1 =& \mathbb{E}\{\omega_x(k)\} = \\ & (1-\beta_{x,k})\mathbb{E}\{\omega_V(k)\} + \beta_{x,k}\mathbb{E}\{\omega_x(k-N)\} \\ & + \beta_{x,k}T_i\left(\sum_{j=k-N}^{k-1}\mathbb{E}\{\omega_{v_x}(j)\} + \sum_{j=k-2N}^{k-N-1}\mathbb{E}\{\omega_{v_x}(j)\}\right) \\ & + \beta_{x,k}\frac{T_i^2}{2}\left(\sum_{j=k-N}^{k-1}\mathbb{E}\{\omega_{a_x}(j)\} + \sum_{j=k-2N}^{k-N-1}\mathbb{E}\{\omega_{a_x}(j)\}\right), \end{aligned} \tag{25}$$

where the evaluation of $\mathbb{E}\{\omega_{v_x}(j)\}$ is presented in Appendix A, while $\mathbb{E}\{\omega_{a_x}(k)\}$ can be computed in a similar way as in the previous subsection. Notice that (25) is similar to (20), and thus the same constraint $\beta_{x,k} \in (-1,1)$ is required.

Finally, P_2 can be derived from (22) by substituting the expression of the error bias (25) in (17), that leads to the following expression for the error variance

$$\begin{aligned} P_2 =& \mathbb{E}\{(\widehat{\delta x}(k) - \mathbb{E}\{\widehat{\delta x}(k)\})^2\} \\ =& \mathbb{E}\left\{((1-\beta_{x,k})v_V(k) + \beta_{x,k}v_x(k-N) + \right. \\ & + \beta_{x,k}\sum_{j=k-N}^{k-1}\left(T_iv_{v_x}(j) + \frac{T_i^2}{2}v_{a_x}(j)\right) + \\ & \left. - \beta_{x,k}\sum_{j=k-2N}^{k-N-1}\left(T_iv_{v_x}(j) + \frac{T_i^2}{2}v_{a_x}(j)\right)\right)^2\Bigg\}. \end{aligned} \tag{26}$$

Since v_{V_x}, v_x, v_{a_x} and v_{v_x} are statistically independent and $\mathbb{E}\{v_{a_x}\} = \mathbb{E}\{v_x\} = 0$, (26) yields

$$\begin{aligned} P_2 =& \sigma^2_x(k) = (1-\beta_{x,k})^2\sigma^2_V(k) + \beta^2_{x,k}\sigma^2_x(k-N) \\ & + \beta^2_{x,k}\left(\sum_{j=k-N}^{k-1}T_i^2\sigma^2_{v_x}(j) + \sum_{j=k-N}^{k-1}\frac{T_i^4}{4}\sigma^2_{a_x}(j)\right) + \\ & - \beta^2_{x,k}\left(\sum_{j=k-2N}^{k-N-1}T_i^2\sigma^2_{v_x}(j) + \sum_{j=k-2N}^{k-N-1}\frac{T_i^4}{4}\sigma^2_{a_x}(j)\right). \end{aligned} \tag{27}$$

IV. Solution of the Pareto Optimization Problem

The optimization problem (18) is convex and can been proven by taking the derivative of the objective function respect to $\beta_{x,k}$. Thus the optimal value of $\beta_{x,k}$ for a fixed $\rho_{x,k}$ is $\beta^*_{x,k} = \max(-1, \min(\xi, 1))$, with

$$\xi = \frac{2(1-\rho_{x,k})\sigma^2_{V_x}(k+1) - 2\rho_{x,k}\gamma_{x,k}\mathbb{E}\{\omega_{V_x}(k+1)\}}{2(1-\rho_{x,k})\eta_{x,k} + 2\rho_{x,k}\gamma^2_{x,k}}.$$

The best value of $\rho_{x,k}$ is found through a Pareto trade-off curve as in Fig. 2, and by selecting the *knee-point* on this curve [23]. Thus, the optimal value $\rho^*_{x,k}$ is chosen such that P_1 and P_2 computed in $\beta^*_{x,k}(\rho^*_{x,k})$ give $P_2 \simeq P_1^2$, that is given by the solution of the following optimization problem

$$\rho^*_{x,k} = \arg\min_{\rho_{x,k}}\left(P_2\left(\beta^*_{x,k}(\rho_{x,k})\right) - P_1^2\left(\beta^*_{x,k}(\rho_{x,k})\right)\right). \tag{28}$$

Being this problem non-linear, a numerical procedures based on the discrimination of $\rho_{x,k}$ can be employed [23].

Notice that the analytical expressions of $\eta_{x,k}$ and $\gamma_{x,k}$ are different for the synchronous and the asynchronous sensor fusion cases. In the case of synchronous measurements, the

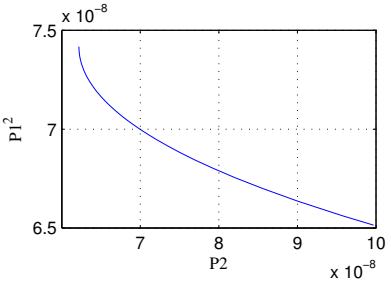


Fig. 2. Pareto tradeoff curve. Each point on the curve is computed considering a different value of $\rho_{x,k}$

expressions of these parameters depend on the estimation of the state bias and variance at the current step time, i.e.

$$\eta_{x,k} = \sigma^2_{V_x}(k) + \sigma^2_x(k-1) + T^4\sigma^2_{a_x}(k-1),$$
$$\gamma_{x,k} = -\mathbb{E}\{\omega_{V_x}(k)\} + \mathbb{E}\{\omega_x(k-1)\} + T^2\mathbb{E}\{\omega_{a_x}(k-1)\}.$$

The solution for the asynchronous case depends on the state bias and variance at current step time, i.e.

$$\eta_{x,k} = \sigma^2_{V_x}(k) + \sigma^2_x(k-1) + T_i^2\left(\sum_{j=k-N}^{k-1}\sigma^2_{v_v}(j) - \sum_{j=k-2N}^{k-N-1}\sigma^2_{v_v}(j)\right) + \frac{T_i^4}{4}\left(\sum_{j=k-N}^{k-1}\sigma^2_{a_x}(j) - \sum_{j=k-2N}^{k-N-1}\sigma^2_{a_x}(j)\right),$$

$$\gamma_{x,k} = -\mathbb{E}\{\omega_{V_x}(k)\} + \mathbb{E}\{\omega_x(k-1)\} + T_i^2\left(\sum_{j=k-N}^{k-1}\mathbb{E}\{\omega_{v_x}(j)\} + \sum_{j=k-2N}^{k-N-1}\mathbb{E}\{\omega_{v_x}(j)\}\right) + \frac{T_i^4}{4}\left(\sum_{j=k-N}^{k-1}\mathbb{E}\{\omega_{a_x}(j)\} - \sum_{j=k-2N}^{k-N-1}\mathbb{E}\{\omega_{a_x}(j)\}\right).$$

Remark 1: The solution to the optimization problem (18) is equivalent to an optimization respect to both $\beta_{x,k}$ and $\rho_{x,k}$ parameters. In fact, by imposing the first-order optimality condition with respect to $\rho_{x,k}$, we obtain $P_2 \simeq P_1^2$.

Then, by imposing the first-order optimality condition with respect to $\beta_{x,k}$, we obtain

$$(1-\rho_{x,k})\frac{\partial P_2}{\partial \beta_{x,k}} + 2\rho_{x,k}P_1\frac{\partial P_1}{\partial \beta_{x,k}} = 0$$

which should be solved numerically by choosing $\beta_{x,k}$ depending on $\rho_{x,k}$ such the first enounced condition $P_2 \simeq P_1^2$ is verified. The Pareto term is derived from the classical game theory where the goal of two players is to choose a given strategy such to maximize (minimize) their utility, in the presented problem represented by P_2 and P_1^2 respectively.

V. Simulations

The proposed method has been tested on simulated trajectories both for the synchronous and the asynchronous case. The considered path is a $3D$ circle generated according to different time law profiles emulating different operating conditions (see Fig. 3).

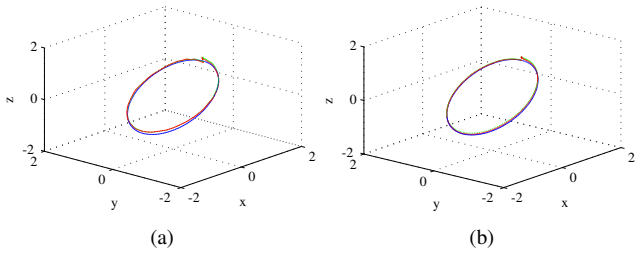
A. System Characterization

The variance of the vision measurement to model a variable distance d with respect to the observed target is

$$\sigma^2_V(k) = \sigma^2_{V\min} + \frac{\sigma^2_{V\max} - \sigma^2_{V\min}}{d_{\max} - d_{\min}}(d(k) - d_{\min}),$$

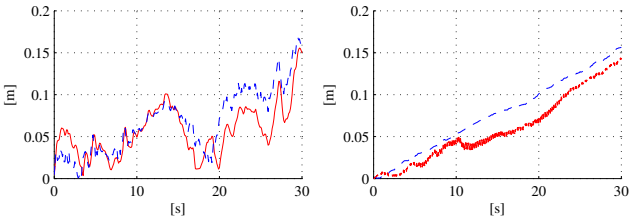
where $d_{\min}$ and $d_{\max}$ represent the minimum and the maximum distance, respectively.

The trajectory estimation error of the proposed optimization technique has been compared with the case when only vision data are employed. In both the synchronous and the asynchronous cases, the estimation error benefits from the proposed approach, as shown in Fig. 4. Tables I and II show the robustness of this method in different working conditions, by considering different time laws.



(a) (b)

Fig. 3. Trajectory paths in the synchronous (*left*) and asynchronous (*right*) case: ground truth (blu), vision based estimation (green), and Pareto estimation (red). *Synchronous case*: $T = 0.1$ s, $\sigma^2_{a_x} = \sigma^2_{a_y} = \sigma^2_{a_z} = 0.2^2$ m^2/s^4, $\sigma_V(d) \in [1, 4]$ mm; pitch rotation $\theta = \pi/6$. *Asynchronous case*: $T_i = 0.01s$, $T_V = 0.1s$, $\sigma^2_{a_x} = \sigma^2_{a_y} = \sigma^2_{a_z} = 0.3^2$ m^2/s^4, $\sigma_V(d) \in [0.1, 0.4]$ mm; roll, pitch and yaw rotation $\phi = -\pi/4, \theta = \pi/8, \psi = -\pi/6$. *Both cases*: $\mathbb{E}\{\omega_{V_x}\} = 0.3$ mm, $\sigma^2_\phi = 0.02^2, \sigma^2_\theta = 0.03^2, \sigma^2_\psi = 0.01^2$ rad/s^2.



(a) Synchronous (b) Asynchronous

Fig. 4. Time history of the norm of the trajectory estimation error with respect to the ground truth by using only vision data (blue dashed line) and with the Pareto optimization (red continuous line).

TABLE I

Average error norm in the synchronous case

Case	Trapez. Velocity	Cubic Poly	5^{th}order Poly
Vision error [m]	0.0676	0.0671	0.0665
Estimation error [m]	0.0568	0.0586	0.0549

TABLE II

Average error norm in the asynchronous case

Case	Trapez. Velocity	Cubic Poly	5^{th}order Poly
Vision error [m]	0.0770	0.0769	0.0760
Estimation error [m]	0.0574	0.0544	0.0497

B. System Performances

The proposed method has been compared to a Stochastic Cloning Kalman filtering (SC-KF) [7] using possible IMU noise values. A second-order dynamic model has been used, where the state is increased by the old visual system position, so as to consider differential visual position measurements.

Differently from the proposed approach, Kalman filtering techniques rely on the state and measurement covariance

matrices, which are typically constant in classic Kalman-filter implementations. The proposed approach, instead, takes into account the variance and the bias on the system state at each instant of time, thus producing a significant benefit.

To compare the two different approaches, the same time-varying law estimated variance, as employed in the proposed method, is used in the SC-KF implementation. Moreover, the bias is modeled as a constant parameter causing a state augmentation. With reference to the synchronous case, Table I show the comparison between the two different mentioned approaches. whereas, Table III and Fig. 5 show the comparison for the asynchronous case. In both cases the proposed problem formulation is able to reduce the error norm by about 30%. In the asynchronous case, an increasing of oscillations in the error norm can be observed, which is caused by the presence of significant noise on acceleration measurements typical on aerial platforms.

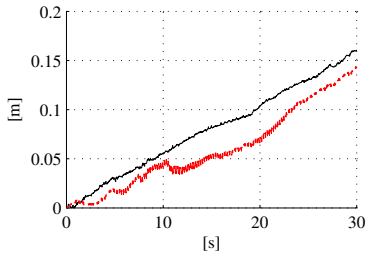


Fig. 5. Asynchronous case: comparison between SC-KF (black point dashed line) and the proposed method (red continuous line).

TABLE III
AVERAGE ERROR NORM

Case	Trapez. Velocity	Cubic Poly	5^{th}order Poly
Proposed method [m]	0.0574	0.0544	0.0497
Sto. Clo. KF [m]	0.0799	0.0798	0.0792

Notice that, when the vision measurement is unavailable, only acceleration is used for the estimation. By starting the estimation from an optimal position value, as shown in (12), it is possible to prevent estimation divergence due to acceleration measurements. The bias $\mathbb{E}\{\omega_{a_x}(k)\}$ and its corresponding variance in case of too noisy acceleration, can be computed from filtered acceleration data obtained using a classical first order low pass filter.

The index (18) minimizes a combination of state bias and variance, while through a KF approach the covariance state matrix is minimized. Being a combination of bias and variance, it can be directly interpreted as a measure of the Mean Square Error (MSE) of the unknown scalar parameter ω_x, similarly to the biased estimation in [24]. This is even confirmed by the analysis of the average covariance values.

The computational complexity of the proposed solution is $\mathrm{O}(n_1 n_2)$ with $n_1 = t/T$, $n_2 = 1/T_\rho$, where t is the current time, $T = T_i$ and T_ρ is the step time used to search the ρ value such that $P_2 \simeq P_1^2$, while for KF approaches it is just $\mathrm{O}(n_1)$. The difference is confirmed by the average computational time of both methods, as shown in Table IV for two different hardware platforms.

TABLE IV
COMPUTATIONAL TIME

Case	Intel Core $i2$	Intel Core $i7$
Proposed method [ms]	2.3	1.6
Sto. Clo. KF [ms]	0.4	0.18

VI. CONCLUSION

In this paper a new sensor fusion technique for motion estimation which combines mono-camera visual and inertial measurements via a Pareto optimization process has been presented. The proposed method minimizes a combination of state bias and variance by balancing available data input in an optimal way. Only the measurements statistical characterization is required, without any prior knowledge of the motion model. The effectiveness of the proposed theoretical approach, in terms of accuracy, computational requirements, and robustness was tested in simulation case studies and compared to a Kalman-filter based approach. It is shown that the proposed method gives a benefit in terms of estimation accuracy with a limited increase of the computational complexity.

Future works will involve the test on a real dataset and an extended analysis on the optimization index.

A. APPENDIX

By developing the third term of (25) we have

$$\sum_{j=k-N}^{k-1} \mathbb{E}\{\omega_{v_x}(j)\} = \mathbb{E}\{\omega_{v_x}(k-N)\} + \mathbb{E}\{\omega_{v_x}(k-N+1)\} + \cdots + \mathbb{E}\{\omega_{v_x}(k-1)\},$$

where every single term can be written as

$$\mathbb{E}\{\omega_{v_x}(k-N+1)\} = \mathbb{E}\{\omega_{v_x}(k-N)\} + T_i\mathbb{E}\{\omega_{a_x}(k-N)\}$$

$$\mathbb{E}\{\omega_{v_x}(k-N+2)\} = \mathbb{E}\{\omega_{v_x}(k-N+1)\} + T_i\mathbb{E}\{\omega_{a_x}(k-N+1)\} = \mathbb{E}\{\omega_{v_x}(k-N)\} + T_i\mathbb{E}\{\omega_{a_x}(k-N)\} + T_i\mathbb{E}\{\omega_{a_x}(k-N+1)\}$$

$$\mathbb{E}\{\omega_{v_x}(k-1))\} = \mathbb{E}\{\omega_{v_x}(k-N)\} + T_i\mathbb{E}\{\omega_{a_x}(k-N)\} + \cdots + T_i\mathbb{E}\{\omega_{a_x}(k-2)\}.$$

Thus it gives

$$\sum_{j=k-N}^{k-1} T_i\mathbb{E}\{\omega_{v_x}(j)\} = (N-1)\mathbb{E}\{\omega_{v_x}(k-N)\} + T_i \sum_{j=k-N}^{k-1} (k-j)\mathbb{E}\{\omega_{a_x}(j)\}.$$

This results shows how the recursive expression of the velocity bias is obtained. The same approach can be used to derive the velocity variance σ_{v_x}.

In the following the symbol σ_i for the variance associated to i will be employed.

Lemma A. 1: Let $\widetilde{\varphi}(k)$ be Gaussian and independent respect to the acceleration component $\widetilde{a}(k)$; then

$$\mathbb{E}\{\widetilde{a}(k)\sin(\widetilde{\varphi}(k))\} = a(k)\sin(\varphi(k))e^{\frac{-\sigma_\varphi^2}{2}}.$$

Proof: Using sine properties, since $\widetilde{a}(k)$ and $\widetilde{\varphi}(k)$ are statistically independent, we obtain that

$$\begin{aligned}&\mathbb{E}\{\widetilde{a}(k)\sin(\widetilde{\varphi}(k))\} = \mathbb{E}\{a(k)\}\mathbb{E}\{\sin(\widetilde{\varphi}(k))\} = \\ &a(k)\mathbb{E}\{\sin(\varphi(k))\}\mathbb{E}\{\cos(\omega_\varphi(k))\}+ \\ &a(k)\mathbb{E}\{\cos(\varphi(k))\}\mathbb{E}\{\sin(\omega_\varphi(k))\}.\end{aligned}$$

As shown in [13], [15]

$$\begin{aligned}\mathbb{E}\{\cos(\omega_\varphi(k))\} &= e^{\frac{-\sigma_\varphi^2}{2}} \\ \mathbb{E}\{\sin(\omega_\varphi(k))\} &= 0,\end{aligned}$$

that leads to

$$\mathbb{E}\{\widetilde{a}(k)\sin(\widetilde{\varphi}(k))\} = a(k)\sin(\varphi(k))e^{\frac{-\sigma_\varphi^2}{2}}.$$

■

Lemma A. 2: Let $\widetilde{\varphi}(k)$ be Gaussian and independent with respect to the acceleration component $\widetilde{a}(k)$; then

$$\mathbb{E}\{\widetilde{a}^2(k)\sin^2(\widetilde{\varphi}(k))\} = \sigma_a^2\left(\frac{1}{2} - \frac{1}{2}\cos(2\varphi)\right)e^{-2\sigma_\varphi^2}.$$

Proof: Since $\widetilde{a}(k)$ and $\widetilde{\varphi}(k)$ are statistically independent, we obtain that

$$\mathbb{E}\{\widetilde{a}^2(k)\sin^2(\widetilde{\varphi}(k))\} = \mathbb{E}\{\widetilde{a}^2(k)\}\mathbb{E}\{\sin^2(\widetilde{\varphi}(k))\}.$$

Then using sine properties yields

$$\begin{aligned}&\mathbb{E}\{\sin^2(\widetilde{\varphi}(k))\} = \mathbb{E}\left\{\frac{1}{2} - \frac{1}{2}\cos(2\widetilde{\varphi}(k))\right\} = \\ &\mathbb{E}\left\{\frac{1}{2} - \frac{1}{2}\cos(2(\varphi(k)+\omega_\varphi(k))\right\} = \\ &\frac{1}{2} - \frac{1}{2}\mathbb{E}\left\{\cos(2(\varphi(k))\cos(2\omega_\varphi(k))+\right. \\ &\left.-\sin(2(\varphi(k))\sin(2\omega_\varphi(k))\right\}.\end{aligned}$$

As shown in [13], [15]

$$\mathbb{E}\{\cos(2\omega_\phi(k))\} = e^{-2\sigma_\varphi^2}.$$

The result can be extended to the $\sin$ term by considering its series expansion

$$\begin{aligned}&\mathbb{E}\{\sin(2\omega_\varphi(k))\} = \mathbb{E}\left\{2\omega_\varphi(k) + \cdots + \right. \\ &\left.+(-1)^n\frac{2\omega_\varphi(k)^{2n+1}}{(2n+1)!}\right\} = 0.\end{aligned}$$

Hence the initial expression becomes

$$\mathbb{E}\{\sin^2(\widetilde{\varphi}(k))\} = \left(\frac{1}{2} - \frac{1}{2}\cos(2\varphi)\right)e^{-2\sigma_\varphi^2}.$$

Then the final result is

$$\mathbb{E}\{\widetilde{a}^2(k)\sin^2(\widetilde{\varphi}(k))\} = \sigma_a^2\left(\frac{1}{2} - \frac{1}{2}\cos(2\varphi)\right)e^{-2\sigma_\varphi^2}.$$

■

REFERENCES

[1] F. Gustafsson, "Statistical sensor fusion", *Studentliteratur*, KTH, 2010.

[2] P. Oguz-Ekim, J. Gomes, J. Xavier, P. Oliveira, "A convex relaxation for approximate maximum likelihood 2d source localization from range measurements", *IEEE Conf. on Acoustics Speech and Signal Processing (ICASSP)*, Dallas, Texas, USA, 2010.

[3] L. Armesto, J. Tornero, M. Vincze, "On multi-rate fusion for non-linear sampled data systems: Application to a 6D tracking system, *Robotics and Autonomous Systems*, 56, 706–715, 2008.

[4] V. Lippiello, B. Siciliano, L. Villani, "Adaptive extended Kalman filtering for visual motion estimation of 3D objects", *Control Engineering Practice*, 15, 123–134, 2007.

[5] T.D. Larsen, N.A. Andersen, O. Ravn, N.K. Poulsen, "Incorporation of time delayed measurements in discrete-time Kalman Filter", *37th IEEE Conf. on Decision and Control*, Tampa, Florida, USA, 1998.

[6] F.H. Hsiao, S.-T. Pan, "Robust Kalman filter synthesis for uncertain multiple time-delay stochastic systems", *ASME Journal of Dynamic Systems, Measurement, and Control*, 118, 803–808, 1996.

[7] A.I. Mourikis, S. Roumeliotis, "On the treatment of relative-pose measurements for mobile robot localization", *2006 IEEE Int. Conf. on Robotics and Automation*, Orlando, Florida, USA, 2006.

[8] L. Kneip, D. Scaramuzza, R. Siegwart, "On the initialization of statistical optimum filters with application to motion estimation", *2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Taipei, Taiwan, 2010.

[9] F. Bourgeois, L. Kneip, S. Weiss, R. Siegwart, "Delay and dropout tolerant state estimation for MAVs", *12th Int. Symposium on Experimental Robotics*, New Delhi and Agra, India, 2010.

[10] L. Kneip, A. Martinelli, S. Weiss, D. Scaramuzza, R. Siegwart, "Closed-form solution for absolute scale velocity determination combining inertial measurements and a single feature correspondence", *2011 IEEE Int. Conf. on Robotics and Automation*, Shanghai, China, 2011.

[11] V. Lippiello, G. Loianno, B. Siciliano, "MAV indoor navigation based on a closed-form solution for absolute scale velocity estimation using optical flow and inertial data", *50th IEEE Conf. on Decison and Control*, Orlando, Florida, USA, 2011.

[12] V. Lippiello, R. Mebarki, "Closed-form solution for absolute scale velocity estimation using visual and inertial data with a sliding least-squares estimation", *21st Mediterranean Conf. on Control and Automation*, Crete, Greece, 2013.

[13] A. De Angelis, C. Fischione, "A distributed information fusion method for localization based on Pareto optimization", *7th IEEE Int. Conf. on Distributed Computing in Sensor Systems*, Barcelona, Spain, 2011.

[14] A. Speranzon, C. Fischione, K. H. Johansson, A. Sangiovanni-Vincentelli, "A distributed information fusion method for localization based on Pareto communications", *IEEE Journal on Selected Areas in Communications*, 26, 609–621, 2008.

[15] A. De Angelis, C. Fischione, P. Händel, "A sensor fusion algorithm for cooperative localization", *IFAC World Congress*, Milan, Italy, 2011.

[16] T. Hamel, R. Mahony "Attitude estimation on $SO(3)$ based on direct inertial measurements", *2006 IEEE Int. Conf. on Robotics and Automation*, Orlando, Florida, 2006.

[17] T. Cheviron, T. Hamel, R. Mahony, G. Baldwin, "Robust nonlinear fusion of inertial and visual data for position, velocity and attitude estimation of UAV", *2007 IEEE Int. Conf. on Robotics and Automation*, Rome, Italy, 2007.

[18] A.J. Baerveldt, R. Klang, "A low-cost and low-weight attitude estimation system for an autonomous helicopter", *1997 IEEE Int. Conf. on Intelligent Engineering Systems*, 391–395, 1997.

[19] S.O.H. Madgwick, A.J.L. Harrison, R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm", *12th IEEE Int. Conf. on Rehabilitation Robotics*, Zürich, Switzerland, 2011.

[20] D. Scaramuzza, F. Fraundorfer, "Visual odometry. Part I: The first 30 years and fundamentals", *IEEE Robotics and Automation Magazine*, 18(4), 80–92, 2011.

[21] F. Fraundorfer, D. Scaramuzza, "Visual odometry – Part II: Matching, robustness, optimization and applications", *IEEE Robotics and Automation Magazine*, 19(2), 78–90, 2012.

[22] A. Georgiev, P.K. Allen, "Localization for mobile robots in urban environments", *IEEE Transactions on Robotics*, 20, 851–864, 2004.

[23] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[24] S. Kay, Y.C. Eldar, "Rethinking biased estimation", *IEEE Signal Processing Magazine*, 25(3), 133–136, 2008.