HAL
open science

Closed-Form Solution to the Structure from Motion Problem by Fusing Visual and Inertial Sensing

Agostino Martinelli

To cite this version:

Agostino Martinelli. Closed-Form Solution to the Structure from Motion Problem by Fusing Visual and Inertial Sensing. 2012. hal-00700231

HAL Id: hal-00700231

https://hal.science/hal-00700231v1

Submitted on 22 May 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Closed-Form Solution to the Structure from Motion Problem by Fusing Visual and Inertial Sensing

Agostino Martinelli

N° ????

Mai 2012

Thèmes COG et NUM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--????--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE
INRIA
centre de recherche GRENOBLE - RHÔNE-ALPES

Closed-Form Solution to the Structure from Motion Problem by Fusing Visual and Inertial Sensing

Agostino Martinelli

Thèmes COG et NUM — Systèmes cognitifs et Systèmes numériques
Équipe-Projet Emotion

Rapport de recherche n° ???? — Mai 2012 — 11 pages

Abstract: The structure from motion problem (SfM) consists in determining the three-dimensional structure of the scene by using the measurements provided by one or more sensors over time (e.g. vision sensors, ego-motion sensors, range sensors). Solving this problem consists in simultaneously performing self motion perception (sMP) and depth perception (DP). In the case of visual measurements only, the SfM has been solved up to a scale [5, 7, 11, 17, 26] and a closed form solution has also been derived [11, 17, 26], allowing the determination of the three-dimensional structure of the scene, without the need for any prior knowledge. The case of inertial and visual measurements has particular interest and has been investigated by many disciplines, both in the framework of computer science [3, 14, 15, 27] and in the framework of neuroscience (the visual-vestibular integration for sMP [2, 10, 18, 30] and for DP [8]). Prior work has answered the question of which are the observable modes, i.e. the states that can be determined by fusing visual and inertial measurements [3, 14, 15, 21, 23]. The questions of how to compute these states in the absence of a prior, and of how many solutions are possible, have only been answered very recently [21, 23]. Here we derive a very simple and intuitive derivation of the solution introduced in [21, 23]. We show that the SfM problem can have a unique solution or two distinct solutions or infinite solutions depending on the trajectory, on the number of point-features and on the number of monocular images where the same point-features are seen. Our results are relevant in all the applications which need to solve the SfM problem with low-cost sensors and which do not demand any infrastructure. Additionally, our results could play an important role in neuroscience by providing a new insight on the process of vestibular and visual integration.

Key-words: Sensor Fusion, Inertial Sensors, Vision, Non linear Observability, Structure from Motion

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Résumé : Le problème *structure from motion* (SfM) consiste en déterminer la structure à trois dimensions de la scène en utilisant les mesures fournies par un ou plusieurs capteurs progressivement (par exemple capteurs optiques, capteurs d'ego-mouvement, capteurs de distance). Résoudre ce problème consiste en résoudre simultanément la perception du mouvement (sMP) et la perception de profondeur (DP). Dans le cas de mesures visuelles seulement, le SfM a été résolu [5, 7, 11, 17, 26] et une solution analytique a été aussi dérivée [11, 17, 26], permettant la détermination de la structure de la scène à moins de l'echelle, sans le besoin d'une connaissance a priori. Le cas de capteurs inertielles et visuelles a un intérêt particulier et a été examiné par beaucoup de disciplines, soit dans le cadre des sciences informatiques [3, 14, 15, 27] soit dans le cadre des neurosciences (l'intégration visuel-vestibulaire pour le sMP [2, 10, 18, 30] et pour le DP [8]). Beaucoup d'effort a été fait pour détecter les modes observables, c.-à-d. les états qui peuvent être déterminés en fusionnant données visuelles et inertielles [3, 14, 15, 21, 23]. La question de comment estimer la valeur de ces états en absence d'une connaissance a priori, et de combien de solutions sont possibles, a été répondu seulement très récemment [21, 23]. Ici nous proposons une dérivation très simple et intuitive de la solution introduite dans [21, 23]. Nous montrons que le problème de SfM peut avoir une solution unique, deux solutions ou encore infinies solutions. Nos résultats sont fondamentales dans toutes les applications qui demandent de résoudre le problème SfM avec des détecteurs à bas prix et qui n'exige pas de l'infrastructure. En plus, nos résultats pourraient jouer un rôle important dans la neuroscience en fournissant une nouvelle prospective sur le processus d'intégration vestibulaire et visuelle.

Mots-clés : Fusion Sensoriel, Capteurs inertiels, Vision, Observabilité, Structure from Motion

1 Introduction

Recent works on visual-inertial navigation have analytically derived the observable states and the set of parameters which are identifiable when fusing monocular vision with inertial sensing [3, 14, 15, 21, 23]. Specifically, it has been shown that the velocity, the absolute scale, the gravity vector in the local frame and the bias-vectors which affect the inertial measurements, are observable modes. On the other hand, the vast majority of existing techniques for navigation using visual and IMU measurements employ either a recursive Bayesian estimation approach [3, 14, 15], or a smoothing formulation [27]. Both types of methods require a prior estimate for the initial state: since both rely on linearization of the nonlinear measurement models, lack of a good initial estimate can lead to divergence. This important limitation would be eliminated by introducing a deterministic solution, i.e., by analytically expressing all the observable modes in terms of the measurements provided by the sensors during a short time-interval. A first derivation of this solution has been introduced in [21, 23]. Here we propose an alternative and more intuitive derivation.

2 Derivation

We consider a given time interval $[t_0,\ T]$. Our goal is to express in closed-form all the observable modes at time t_0 only in terms of the visual and inertial measurements obtained during $[t_0,\ T]$. In order to characterize the motion of a body moving in a $3D$ environment we introduce a global frame. Its z-axis points vertically upwards. As we will see, for the next derivation we do not need to better define this global frame. We will adopt lower-case letters to denote vectors in this frame (e.g. the gravity is $\boldsymbol{g} = [0,\ 0,\ -g]^T$, where $g \simeq 9.8\ ms^{-2}$). We also introduce a local frame on the body and we will adopt upper-case letters to denote vectors in this frame. Since this local frame is time dependent, we adopt the following notation: $\boldsymbol{W}_t(\tau)$ will be the vector with global coordinates $\boldsymbol{w}(\tau)$ in the local frame at time t. Additionally, we will denote with $C_{t_1}^{t_2}$ the matrix which characterizes the rotation occurred during the time interval $(t_1,\ t_2)$. We have: $\boldsymbol{W}_{t_1}(\tau) = C_{t_1}^{t_2}\boldsymbol{W}_{t_2}(\tau)$ and $(C_{t_1}^{t_2})^{-1} = C_{t_2}^{t_1}$. Finally, C^t will denote the rotation matrix between the global frame and the local frame at time t, i.e., $\boldsymbol{w}(\tau) = C^t\boldsymbol{W}_t(\tau)$.

The position of the body $\boldsymbol{r}$ at any time $t \in [t_0,\ T]$ satisfies the following equation:

$$\boldsymbol{r}(t) = \boldsymbol{r}(t_0) + \boldsymbol{v}(t_0)\Delta t + \boldsymbol{g}\frac{\Delta t^2}{2} + C^{t_0}\boldsymbol{S}_{t_0}(t) \tag{1}$$

where $\boldsymbol{v} \equiv \frac{d\boldsymbol{r}}{dt}$, $\Delta t \equiv t - t_0$, $\boldsymbol{S}_{t_0}(t) \equiv \int_{t_0}^{t}\int_{t_0}^{\tau} C_{t_0}^{\xi}\boldsymbol{A}_{\xi}(\xi)d\xi d\tau$. Note that the matrix $C_{t_0}^{\xi}$ can be obtained from the angular speed during the interval $[t_0,\ \xi]$ provided by the gyroscopes [9] and $\boldsymbol{A}_{\xi}(\xi)$ is the acceleration provided by the accelerometers [9]. In other words, $C_{t_0}^{\xi}\boldsymbol{A}_{\xi}(\xi)$ is the acceleration of the vehicle at the time $\xi \in (t_0,\ T)$ expressed in the vehicle frame at time t_0. The vector $\boldsymbol{S}_{t_0}(t)$ can be obtained at any time by integrating the data provided by the gyroscopes and the accelerometers delivered during the interval $[t_0,\ \xi]$.

Let us suppose that N_f point-features are observed, simultaneously. Let us denote their position in the physical world with $\boldsymbol{p}^i,\ i = 1,\ ...,\ N_f$. According

to our notation, $\boldsymbol{P}_t^i(t)$ will denote their position at time t in the local frame at time t. We have: $\boldsymbol{p}^i = \boldsymbol{r}(t) + C^{t_0} C_{t_0}^t \boldsymbol{P}_t^i(t)$ and, by using (1), we obtain the following fundamental equation:

$$\boldsymbol{P}_t^i(t) = C_t^{t_0} \left[\boldsymbol{P}_{t_0}^i(t_0) - \boldsymbol{V}_{t_0}(t_0)\Delta t - \boldsymbol{G}_{t_0}(t_0)\frac{\Delta t^2}{2} - \boldsymbol{S}_{t_0}(t) \right] \tag{2}$$

Our goal is to determine the observable modes, i.e., $\boldsymbol{P}_{t_0}^i(t_0)$ $(i = 1, \ ..., \ N_f)$, $\boldsymbol{V}_{t_0}(t_0)$ and $\boldsymbol{G}_{t_0}(t_0)$. We remark that the determination of $\boldsymbol{G}_{t_0}(t_0)$ is equivalent to the knowledge of the roll and pitch angles at the time t_0 together with the magnitude of the gravity (g). Let us denote by$\boldsymbol{X}$ the $(3N_f + 6)$-column vector whose entries are the unknowns $(\boldsymbol{X} \equiv [\boldsymbol{P^1}_{t_0}(t_0), \ \boldsymbol{P^2}_{t_0}(t_0), \ ..., \ \boldsymbol{P^{N_f}}_{t_0}(t_0), \ \boldsymbol{V}_{t_0}(t_0), \ \boldsymbol{G}_{t_0}(t_0)]^T)$. Equation (2) states that the position of any point-feature in the local frame at time t is affine in $\boldsymbol{X}$. Let us suppose that the N_f point-features are observed in n_{im} consecutive images. A single image provides the bearing angles of all the point-features in the local frame. In other words, an image taken at the time t provides all the vectors $\boldsymbol{P}_t^i(t)$ up to a scale. Since the data provided by the gyroscopes during the interval $(t_0, \ T)$ allow us to build the matrix $C_{t_0}^t$, having the vectors $\boldsymbol{P}_t^i(t)$ up to a scale, allows us to also know the vectors $\boldsymbol{P}_{t_0}^i(t) = C_{t_0}^t \boldsymbol{P}_t^i(t)$ up to a scale. From equation (2) we easily obtain:

$$\boldsymbol{P}_{t_0}^i(t) = \boldsymbol{P}_{t_0}^i(t_0) - \boldsymbol{V}_{t_0}(t_0)\Delta t - \boldsymbol{G}_{t_0}(t_0)\frac{\Delta t^2}{2} - \boldsymbol{S}_{t_0}(t) \tag{3}$$

namely, also the vector $\boldsymbol{P}_{t_0}^i(t)$, which can be obtained up to a scale by using the camera and gyroscope data in the interval $(t_0, \ T)$, is affine in $\boldsymbol{X}$.

Knowing a vector up to a scale means to know the ratios among its components. Therefore, the jth image provides, for any point-feature $(i = 1, \ ..., \ N_f)$, the two ratios $y_1^{ij} = \frac{[\boldsymbol{P}_{t_0}^i(t_j)]_x}{[\boldsymbol{P}_{t_0}^i(t_j)]_z}$ and $y_2^{ij} = \frac{[\boldsymbol{P}_{t_0}^i(t_j)]_y}{[\boldsymbol{P}_{t_0}^i(t_j)]_z}$, where $t_1 < t_2 < .... < t_{n_{im}}$ are the times when the n_{im} camera images are taken. Hence we have:

$$[\boldsymbol{P}_{t_0}^i(t_j)]_x = [\boldsymbol{P}_{t_0}^i(t_j)]_z y_1^{ij} \qquad [\boldsymbol{P}_{t_0}^i(t_j)]_y = [\boldsymbol{P}_{t_0}^i(t_j)]_z y_2^{ij} \tag{4}$$

By using (3) it is easy to realize that these are $2n_{im}N_f$ linear equations in the unknowns. In other words, we obtain the following linear system:

$$\alpha \boldsymbol{X} = \boldsymbol{\beta} \tag{5}$$

where α is a $2n_{im}N_f \times (3N_f + 6)$ matrix and $\boldsymbol{\beta}$ is a $(2n_{im}N_f)$-column vector. The entries of α and $\boldsymbol{\beta}$ are directly obtained from the bearing measurements obtained from the images and the gyroscopes and the accelerometers measurements during the interval $(t_0, \ T)$.

3 Number of solutions

In order to understand whether the problem can be solved or not it is necessary to compute the rank of α. In particular, we are interested in understanding how this rank depends on the motion, on the number of observed point-features and on the number of camera images. We have performed an observability analysis [21, 23] based on the rank criterion [12, 13, 22]. The computation was carried

out by using the symbolic computational tool of matlab in order to obtain the expression of the Lie derivatives up to the 4th order together with their gradients. We have extended the results obtained in literature [14, 15] by proving that the observability of the modes holds for any number of point-features (in particular, it holds even in the case of one point). More importantly for our purposes here, we proved that:

1. the observability of all the modes requires the body to move with a varying acceleration and to observe the same point-features in at least four images;
2. when the body moves with constant acceleration all the modes are observable with the exception of the magnitude of the gravity;
3. when the magnitude of the gravity is a priori known and the vehicle moves with constant speed all the remaining modes are observable up to a scale;
4. in general, by only having three images, all the modes are observable with the exception of the magnitude of the gravity.

From these results, we obtain that α cannot be full rank when $n_{im} \leq 3$ and/or when the body moves with constant acceleration. On the other hand, the cases $n_{im} = 3,\ N_f \geq 2$ and the case when the vehicle moves with a constant (non null) acceleration, can still be solved in closed form by exploiting an a priori knowledge of the magnitude of the gravity (which is the unique non-observable mode for these cases). In these cases, since only one mode (g) is unobservable, the matrix α has in general a null space of dimension one (we found special cases where the null space has dimension larger than one, e.g., when the point-features and the body positions where the images are taken are coplanar [21]). Let us denote by $\boldsymbol{n}$ the unit vector spanning this null space. The linear system $\alpha \boldsymbol{X} = \boldsymbol{\beta}$ has infinite solutions. Each solution satisfies the following equation: $\boldsymbol{X} = \alpha^* \boldsymbol{\beta} + \gamma \boldsymbol{n}$, being α^* the pseudoinverse of α and γ a scalar number [24]. The determination of γ is obtained by enforcing the constraint that the norm of the vector formed by the last three elements of $\boldsymbol{X}$ is equal to g.

$$|\P\,(\alpha^* \boldsymbol{\beta} + \gamma \boldsymbol{n})| = g, \quad \P \equiv [0_{3\times(3N_f+3)},\ I_3] \tag{6}$$

where $0_{n\times m}$ is the $n \times m$ matrix whose entries are all zero and I_3 is the identity 3×3 matrix. The equation in (6) is a quadratic polynomial in γ and has two real roots. Hence, we obtain in general two distinct solutions for $\boldsymbol{X}$.

When $n_{im} \geq 4$ and for a motion with a varying acceleration, the matrix α is full rank apart the case $n_{im} = 4,\ N_f = 1$ (actually, there are special cases where α is not full rank, e.g., when the point-features and the four body positions are coplanar [21]). In this latter case the number of equations is 8 and the number of unknowns 9. As in the cases $n_{im} = 3,\ N_f \geq 2$ the problem has in general two distinct solutions which can be determined by solving (6) where $\P \equiv [0_{3\times 6},\ I_3]$.

Table 1 summarizes these results by providing the number of solutions case by case. Figures 1 and 2 illustrate two cases when the SfM problem has two distinct solutions. The body configurations and the position of the point-features in the global frame are shown. The two solutions are in blue and red. The body configuration at the initial time is the same for both the solutions and is in black. Fig 1 regards the case of one feature seen in four images. Fig 2 displays the case of constant acceleration: the case of three point-features in

Cases	Number of Solutions
Varying Acceleration Point-features and motion non-coplanar $n_{im} = 4,\ N_f \geq 2$; $n_{im} \geq 5,\ \forall N_f$	Unique Solution
Varying Acceleration Point-features and motion non-coplanar $n_{im} = 3,\ N_f \geq 2$; $n_{im} = 4,\ N_f = 1$	Two Solutions
Constant and non null Acceleration Point-features and motion non-coplanar $n_{im} = 3,\ N_f \geq 2$; $n_{im} \geq 4,\ \forall N_f$	Two Solutions
Null Acceleration Any Point-features layout $\forall n_{im},\ \forall N_f$	Infinite Solutions
Any Motion Any Point-features layout $n_{im} \leq 2,\ \forall N_f$; $n_{im} = 3,\ N_f = 1$	Infinite Solutions

Table 1: Number of distinct solutions for the SfM problem with visual and inertial sensing

seven images has been considered and the seven poses of the body at the time when the images are taken is shown in the figure together with the position of the features.

4 Extension to the biased case

The closed-form solution here derived can be easily extended to the case when the accelerometers measurements are biased. In this case, equation (2) becomes:

$$\boldsymbol{P}^i_{t_0}(t) = \boldsymbol{P}^i_{t_0}(t_0) - \boldsymbol{V}_{t_0}(t_0)\Delta t - \boldsymbol{G}_{t_0}(t_0)\frac{\Delta t^2}{2} - \int_{t_0}^{t}\int_{t_0}^{\tau} C^{\xi}_{t_0} d\xi d\tau \boldsymbol{A_{bias}} - \boldsymbol{S}_{t_0}(t) \quad (7)$$

and the same procedure previously described can be adopted by considering the extended state $\boldsymbol{X_E} \equiv [\boldsymbol{X},\ \boldsymbol{A_{bias}}]^T$, which includes the three accelerometer bias-components.

5 Discussion and conclusion

The most useful applications of the closed-form solution here derived will be in all the applicative domains which need to solve the SfM problem with low-cost sensors and which do not demand any infrastructure (e.g., in GPS denied environment). In these contexts, there is often the need to perform the estimation without any prior knowledge. Typical examples of applicative domains are the emergent fields of space robotics [25], humanoid robotics and unmanned aerial navigation in urban-like environments [28], where the use of the GPS is often forbidden.

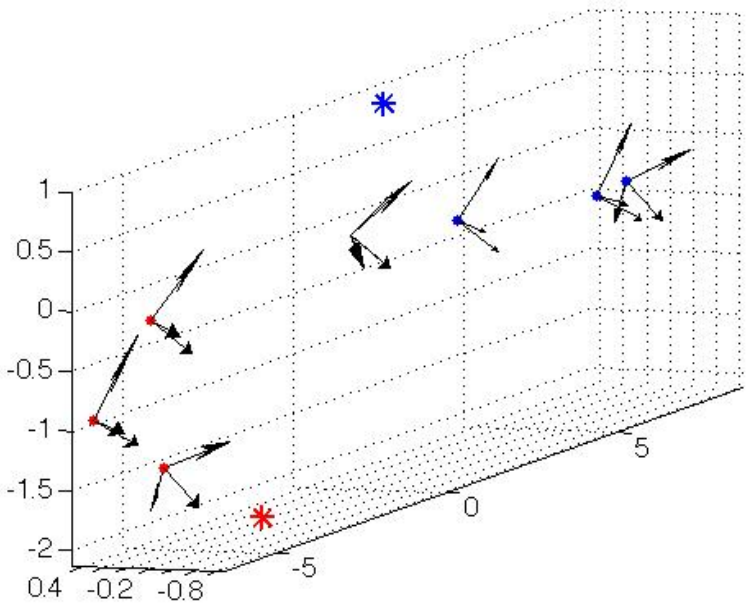

Figure 1: Illustration of two distinct solutions for the case $n_{im} = 4,\ N_f = 1$ (star symbols indicate the position of the point-feature respectively for the two solutions).

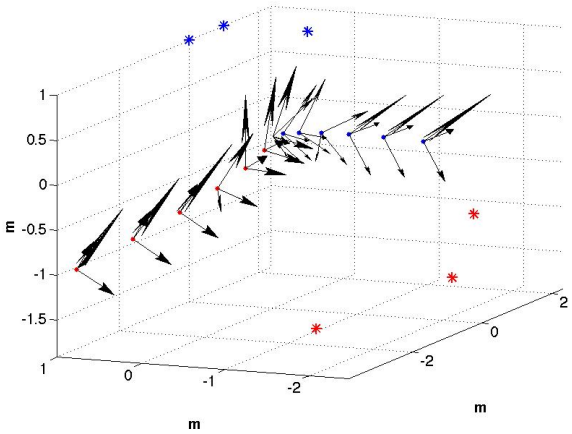

Figure 2: Illustration of two distinct solutions for the case of constant acceleration (star symbols indicate the position of the point-features respectively for the two solutions).

The closed-form solution could also play an important role in the framework of neuroscience by providing a new insight on the process of vestibular and visual integration for DP and sMP. The influence of extra retinal cues in depth perception has extensively been investigated in the last decades [16, 19, 6, 29]. In particular, a very recent study investigates this problem by performing trials with passive head movements [8]. The conclusion of this study is that the combination of retinal image with vestibular signals can provide rudimentary ability to depth perception. Our findings could provide a new insight to this integration mechanism for depth perception since, according to the closed-solution here derived, by combining retinal image with vestibular signals it is possible to determine the scale factor even without any knowledge about the initial speed. Our findings also show that it is possible to easily distinguish linear acceleration from gravity. Specifically, the closed form solution performs this determination by a very simple matrix inversion. This problem has also been investigated in neuroscience [1, 20]. Our results could provide a new insight to this mechanism since they clearly characterize the conditions (type of motion, features layout) under which this determination can be performed.

References

[1] Angelaki D. E., Shaikh A. G., Green A. M. and Dickman J. D., Neurons compute internal models of the physical laws of motion, Nature, 430, pp 560–564, 2004

[2] A. Berthoz, B. Pavard and L.R. Young, Perception of Linear Horizontal Self-Motion Induced by Peripheral Vision (Linearvection) Basic Characteristics and Visual-Vestibular Interactions, Exp. Brain Res. 23, 471–489 (1975).

[3] M. Bryson and S. Sukkarieh, Observability Analysis and Active Control for Airbone SLAM, IEEE Transaction on Aerospace and Electronic Systems, vol. 44, no. 1, 261–280, 2008

[4] A. B. Chatfield, Fundamentals of High Accuracy Inertial Navigation. Reston, VA: American Institute of Aeronautics and Astronautics, Inc., 1997.

[5] Alessandro Chiuso, Paolo Favaro, Hailin Jin and Stefano Soatto, "Structure from Motion Causally Integrated Over Time", IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4), pp 523–535, 2002

[6] Cornilleau-Peres V. and Droulez J., The visual perception of three-dimensional shape from self-motion and object-motion, Vision research, 34, pp 2331–2336, 1994

[7] Andrew J. Davison, Ian D. Reid, Nicholas D.Molton and Olivier Stasse, "MonoSLAM: Real-Time Single Camera SLAM", IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6), pp 1052–1067, 2007

[8] Dokka K., MacNeilage P. R., De Angelis G. C. and Angelaki D. E., Estimating distance during self-motion: a role for visual-vestibular interactions, Journal of Vision (2011) 11(13):2, 1-16

[9] J. A. Farrell, Aided Navigation: GPS and High Rate Sensors. McGraw- Hill, 2008.

[10] C. R. Fetsch, G. C. DeAngelis and D. E. Angelaki, Visual-vestibular cue integration for heading perception: Applications of optimal cue integration theory, Eur J Neurosci. 2010 May ; 31(10): 1721-1729

[11] Richard I. Hartley (June 1997). "In Defense of the Eight-Point Algorithm". IEEE Transaction on Pattern Recognition and Machine Intelligence 19 (6): 580–593.

[12] Hermann R. and Krener A.J., 1977, Nonlinear Controllability and Observability, Transaction On Automatic Control, AC-22(5): 728-740

[13] Isidori A., Nonlinear Control Systems, 3rd ed., Springer Verlag, 1995.

[14] E. Jones and S. Soatto, "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach", The International Journal of Robotics Research, vol. 30, no. 4, pp. 407–430, Apr. 2011.

[15] J. Kelly and G. Sukhatme, Visual-inertial simultaneous localization, mapping and sensor-to-sensor self-calibration, Int. Journal of Robotics Research, vol. 30, no. 1, pp. 56–79, 2011.

[16] Longuet-Higgins H. C. and Prazdny K., The interpretation of a moving retinal image, Royal Society of London B: Biological Sciences, 208, pp 385–397, 1980

[17] H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". Nature 293: 133–135.

[18] P. R. MacNeilage, N. Ganesan and D. E. Angelaki, Computational Approaches to Spatial Orientation: From Transfer Functions to Dynamic Bayesian Inference, J Neurophysiol 100: 2981-2996, 2008

[19] MacNeilage P. R., Banks M. S. De Angelis G. C. and Angelaki D. E, Vestibular heading discrimination and sensitivity to linear acceleration in head and world coordinates, J. of Neuroscience, 30, pp 9084–9094, 2010

[20] Merfeld D. M., Zupan L. and Peterka R. J., Humans use internal models to estimate gravity and linear acceleration, Nature, 398, pp 615–618, 1999

[21] A. Martinelli, Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination by Fusing Vision and Inertial Measurements, Internal Research Report, 2011, INRIA, http://hal.inria.fr/inria-00569083/en/

[22] A. Martinelli, State Estimation Based on the Concept of Continuous Symmetry and Observability Analysis: the Case of Calibration, Transactions on Robotics, Vol. 27, No. 2, pp 239–255, April 2011

[23] A. Martinelli, Vision and IMU Data Fusion: Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination, Transaction on Robotics, Volume 28 (2012), Issue 1 (February), pp 44–60.

[24] C. D. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, 2000

[25] A.I. Mourikis, N. Trawny, S.I. Roumeliotis, A. Johnson, A. Ansar, and L. Matthies, "Vision-Aided Inertial Navigation for Spacecraft Entry, Descent, and Landing", Transactions on Robotics, 25(2), pp. 264-280

[26] D. Nistér, An efficient solution to the five-point relative pose problem, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 26(6):756-770, June 2004

[27] D. Strelow and S. Singh, Motion estimation from image and inertial measurements, International Journal of Robotics Research, 23(12), 2004

[28] Weiss., S., Scaramuzza, D., Siegwart, R., Monocular-SLAM-Based Navigation for Autonomous Micro Helicopters in GPS-Denied Environments, Journal of Field Robotics, Volume 28, issue 6, 2011

[29] Wexler M., Panerai F., Lamouret I. and Droulez J., Self-motion and perception of stationary objects, Nature, 409, pp 85–88, 2001

[30] L. H. Zupan, D. M. Merfeld, C. Darlot, Using sensory weighting to model the influence of canal, otolith and visual cues on spatial orientation and eye movements, Biol. Cybern. 86, 209-230 (2002)

Contents

1 Introduction **3**

2 Derivation **3**

3 Number of solutions **4**

4 Extension to the biased case **6**

5 Discussion and conclusion **6**

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399