Control and estimation algorithms for the stabilization of VTOL UAVs from mono-camera measurements

H. de Plinval[1], A. Eudes[2], P. Morin[2]

[1] ONERA-DCSD, Toulouse, France, henry.de_plinval@onera.fr
[2] Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie
CNRS, UMR 7222
Paris, France, surname@isir.upmc.fr

June 17, 2014

Abstract

This paper concerns the control of Vertical Take-Off and Landing (VTOL) Unmanned Aerial Vechicles (UAVs) based on exteroceptive measurements obtained from a mono-camera vision system. By assuming the existence of a locally planar structure in the field of view of the UAV's camera, the so-called *homography matrix* can be used to represent the vehicle's motion between two views of the structure. In this paper we report recent results on both the problem of homography estimation from the fusion of visual and inertial data, and the problem of feedback stabilization of VTOL UAVs from homography measurements.

1 Introduction

Obtaining a precise estimation of the vehicle's position is a major issue in aerial robotics. The GPS is a very popular sensor in this context and it has been used extensively with VTOL UAVs, expecially for navigation via waypoints. Despite recent progress of this technology, especially in term of precision, many applications cannot be addressed with the GPS as unique position sensor. First, GPS is not available indoor and it can also be masked in some outdoor environments. Then, most inspection applications require a *relative* localization with respect to (w.r.t.) the environment, rather than an absolute localization as provided by the GPS. Finally, evolving in dynamic environments also requires relative localization capabilities. For all these reasons, it is important to develop control strategies based on exteroceptive sensors that can provide a relative position information w.r.t the local environement. Examples of such sensors are provided by cameras, lasers, radars, etc. Cameras are interesting sensors to use with small UAVs because they are light, low cost, and provide a rich information about the environment at a relatively high frequency. A precise 3D relative position information is best obtained from a *stereo* vision system with a "long" baseline (i.e. interdistance between the optical centers of the cameras). In this case, available feedback controllers that require position errors as inputs can be used. Using a mono-camera system is more challenging because the depth-information cannot be recovered instantaneously (i.e., based on a single measurement). Nevertheless, a mono-camera system can be preferred in some applications due to its compacity, or because the distance between the camera and the environment is large so that even a stereo-system would provide a poor depth-information.

This paper concerns the control of VTOL UAVs from mono-camera measurements. We assume the existence of a locally planar structure in the environe-

ment. This assumption is restrictive but it is relevant in practice because *i)* many man-made buildings are locally planar, and *ii)* when the distance between the camera and the environment is large, the planarity assumption can be satisfied locally in first approximation despite the environment not being perfectly planar (e.g. as in the case of ground observation at relatively high altitude). Based on two camera views of this planar structure, it is well known in computer vision that one can compute the so-called *homography matrix*, which embeds all the displacement information between these two views [15]. This matrix can be estimated without any specific knowledge on the planar structure (like its size or orientation). Therefore, it is suitable for the control of UAVs operating in unknown environments. Homography-based stabilization of VTOL UAVs raises two important issues. The first one is the estimation of the homography matrix itself. Several algorithms have been developed in the computer vision community to obtain such an estimation (see, e.g., [15, 1]). Recently, IMU-aided fusion algorithms have been proposed to cope with noise and robustness limitations associated with homography estimation algorithms based on vision data only [16, 9]. The second issue concerns the design of stabilizing feedback laws. The homography associated with two views of a planar scene is directly related to the cartesian displacement (in both position and orientation) between these two views but this relation depends on unknown parameters (normal and distance to the scene). Such uncertainties significantly complicate the design and stability analysis of feedback controllers. This is all the more true that VTOL UAVs are usually underactuated systems, with high-order dynamic relations between the vehicle's position and the control input For example, horizontal displacement is related to roll and pitch control torque via fourth-order systems. For this reason, most existing control strategies based on homography measurements make additional assumptions on the environment, i.e. the knowledge of the normal to the planar scene [20, 21, 18, 14]. This simplifies the control design and stability analysis since in this case, the vehicle's cartesian displacement (rotation and position up to an unknown scale factor) can be extracted from the homography measurement.

This paper reports recent results by the authors and co-authors on both the problem of homography estimation via the fusion of inertial and vision data [16, 9], and the design of feedback controllers based on homography measurements [5, 7]. The paper is organized as follows. Preliminary background and notation are given in section 2. Feedback control algorithms are presented in section 3 and homography estimation algorithms in section 4. Finally, some implementation issues are discussed in section 5.

2 Background

In this section we review background on both the dynamics of VTOL UAVs and the homography matrix associated with two camera images of a planar scene. Let us start by defining the control problem addressed in this paper.

2.1 Control problem

Figure 1 illustrates the visual servoing problem addressed in this paper. A VTOL UAV is equipped with of a mono-camera. A *reference image* of a planar scene $\mathcal{T}$, which was obtained with the UAV located at a reference frame $\mathcal{R}^*$, is available. From this reference image and the current image, obtained from the current UAV location (frame $\mathcal{R}$), the objective is to design a control law that can asymptotically stabilize $\mathcal{R}$ to $\mathcal{R}^*$. Note that asymptotic stabilization is possible only if $\mathcal{R}^*$ corresponds to a possible equilibrium, i.e., in the absence of wind the thrust direction associated with $\mathcal{R}^*$ must be vertical.

2.2 Dynamics of VTOL UAVs

We consider the class of thrust-propelled underactuated vehicles consisting of rigid bodies moving in $3D$-space under the action of one body-fixed force control and full torque actuation [13]. This class contains most VTOL UAVs (quadrotors, ducted fans, helicopters, etc). Being essentially interested here in hovering stabilization, throughout the paper we neglect aerodynamic forces acting on the vehicle's main

body. Assuming that $\mathcal{R}^*$ is a NED (North-East-Down) frame (See Fig. 1), the dynamics of these systems is described by the following well-known equations:

$$\left\{ \begin{array}{lcl} m\ddot{p} & = & -TRb_3 + mgb_3 \\ \dot{R} & = & RS(\omega) \\ J\dot{\omega} & = & J\omega \times \omega + \Gamma \end{array} \right. \quad (1)$$

with p the position vector of the vehicle's center of mass, expressed in $\mathcal{R}^*$, R the rotation matrix from $\mathcal{R}$ to $\mathcal{R}^*$, ω the angular velocity vector of $\mathcal{R}$ w.r.t. $\mathcal{R}^*$ expressed in $\mathcal{R}$, $S(.)$ the matrix-valued function associated with the cross product, i.e. $S(x)y = x \times y$, $\forall x, y \in \mathbb{R}^3$, m the mass, T the thrust control input, $b_3 = (0,0,1)^T$, J the inertia matrix, Γ the torque control input, and g the gravity constant.

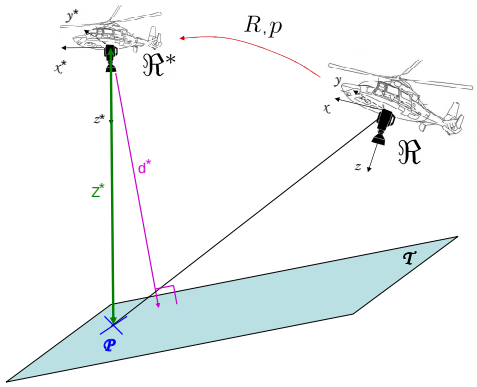

Figure 1: Problem scheme

2.3 Homography matrix and monocular vision

With the notation of Figure 1, consider a point $\mathcal{P} \in \mathcal{T}$ and denote by X^* the coordinates of this point in $\mathcal{R}^*$. In $\mathcal{R}^*$, the plane $\mathcal{T}$ is defined as $\{X^* \in \mathbb{R}^3 : n^{*T}X^* = d^*\}$ with n^* the coordinates in $\mathcal{R}^*$ of the unit vector normal to $\mathcal{T}$ and d^* the distance between the origin of $\mathcal{R}^*$ and the plane. Let us now denote as X the coordinates of $\mathcal{P}$ in the current frame $\mathcal{R}$. One has $X^* = RX + p$ and therefore,

$$\begin{array}{rcl} X & = & R^TX^* - R^Tp \\ X & = & R^TX^* - R^Tp[\frac{1}{d^*}n^{*T}X^*] \\ & = & (R^T - \frac{1}{d^*}R^Tpn^{*T})X^* \\ & = & \bar{H}X^* \end{array} \quad (2)$$

with

$$\bar{H} = R^T - \frac{1}{d^*}R^Tpn^{*T} \quad (3)$$

The matrix $\bar{H}$ could be determined by matching 3D-coordinates in the reference and current camera planes of points of the planar scene. Camera do not provide these 3D-coordinates, however, since only the 2D-projective coordinates of $\mathcal{P}$ on the respective image planes are available. More precisely, the 2D-projective coordinates of $\mathcal{P}$ in the reference and current camera planes are respectively given by

$$\mu^* = K\frac{X^*}{z^*}, \quad \mu = K\frac{X}{z}$$

where z^* and z denote the third coordinate of X^* and X respectively (i.e., the coordinate along the camera optical axis), and K is the calibration matrix of the camera. It follows from (2) and (4) that

$$\mu = G\mu^* \quad (4)$$

with

$$G \propto K\bar{H}K^{-1}$$

where $\propto$ denotes equality up to a positive scalar factor. The matrix $G \in \mathbb{R}^{3\times3}$, defined up to a scale factor, is called *uncalibrated homography matrix*. It can be computed by matching *projections* onto the reference and current camera planes of points of the planar scene. If the camera calibration matrix K is known, then the matrix $\bar{H}$ can be deduced from G, up to a scale factor, i.e., $K^{-1}GK = \alpha\bar{H}$. As a matter of fact, the scale factor α corresponds to the mean singular value of the matrix $K^{-1}GK$: $\alpha = \sigma_2(K^{-1}GK)$ (see, e.g., [15, Pg. 135]). Therefore, α can be computed together with the matrix $\bar{H}$. Another interesting matrix is

$$H = \det(\bar{H})^{-\frac{1}{3}}\,\bar{H} = \eta\,\bar{H} \quad (5)$$

Indeed, $\det(H) = 1$ so that H belongs to the Special Linear Group $SL(3)$. We will see further that this property can be exploited for homography filtering and estimation purposes. Let us finally remark that $\eta^3 = \frac{d^*}{d}$.

3 Feedback Control Design

We present in this section two classes of feedback control laws for the asymptotic stabilization of VTOL UAVs based on homography measurements of the form $\bar{H}$ defined by (3). The first class consists of control laws affine w.r.t. the homography matrix components. These control laws ensure local asymptotic stabilization under very mild assumptions on the observed scene. The second class consists of nonlinear control laws that ensure large stability domains under stronger assumptions on the scene.

3.1 Linear control laws

The main difficulty in homography-based stabilization comes from the mixing of position and orientation information in the homography matrix components, as shown by relation (3). If the normal vector n^* is known, then one can easily extract from $\bar{H}$ the rotation matrix and the position vector up to the scale factor $1/d^*$. When n^* is unknown, however, this extraction is no longer possible and one has to deal with this mixing of information. The control laws here presented rely on the possibility of extracting from $\bar{H}$ *partially decoupled* position and rotation information. This is shown by the following result first proposed in [6].

Proposition 1 *Let $\bar{e} = Me$ with*

$$M = \begin{pmatrix} 2I_3 & S(m^*) \\ -S(m^*) & I_3 \end{pmatrix}, \quad e = \begin{pmatrix} e_p \\ e_\Theta \end{pmatrix} \tag{6}$$

and

$$\begin{gathered} e_p = (I - \bar{H})m^* \ , \ e_\Theta = \mathrm{vex}(\bar{H}^T - \bar{H}) \\ m^* = b_3 = (0,0,1)^T \end{gathered} \tag{7}$$

where vex(.) *is the inverse of the $S(.)$ operator:* $\mathrm{vex}(S(x)) = x \ , \ \forall x \in \mathbb{R}^3$. *Let $\Theta = (\phi, \theta, \psi)^T$ denote any parametrization of the rotation matrix R such that $R \approx I_3 + S(\Theta)$ around $R = I_3$ (e.g., Euler angles). Then,*

1. *$(p, R) \longmapsto \bar{e}$ defines a local diffeomorphism around $(p, R) = (0, I_3)$. In particular, $\bar{e} = 0$ if and only if $(p, R) = (0, I_3)$.*
2. *In a neighborhood of $(p, R) = (0, I_3)$,*

$$\bar{e} = L \begin{pmatrix} p \\ \Theta \end{pmatrix} + O^2(p, \Theta), \quad L = \begin{pmatrix} L_p & 0 \\ L_{p\Theta} & L_\Theta \end{pmatrix} \tag{8}$$

with $L_{p\Theta} = S((\alpha^*, \beta^*, 0)^T)$,

$$L_p = \begin{pmatrix} c^* & 0 & \alpha^* \\ 0 & c^* & \beta^* \\ 0 & 0 & 2c^* \end{pmatrix}, L_\Theta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

α^*, β^* the (unknown) constant scalars defined by $n^* = d^*(\alpha^*, \beta^*, c^*)^T$, $c^* = \frac{1}{\|X^*\|}$, and O^2 terms of order two at least. △

Eq. (8) shows the rationale behind the definition of $\bar{e}$: at first order, components $\bar{e}_1, \bar{e}_2, \bar{e}_3$ contain information on the translation vector p only, while components $\bar{e}_4, \bar{e}_5, \bar{e}_6$ contain decoupled information on the orientation (i.e. L_Θ is diagonal), corrupted by components of the translation vector. Although the decoupling of position and orientation information in the components of $\bar{e}$ is not complete, it is sufficient to define asymptotically stabilizing control laws, as shown below.

Let $\bar{e}_p \in \mathbb{R}^3$ (resp. $\bar{e}_\Theta \in \mathbb{R}^3$) denote the first (resp. last) three components of $\bar{e}$, i.e. $\bar{e} = (\bar{e}_p^T, \bar{e}_\Theta^T)^T$. The control design relies on a dynamic extension of the state vector defined as follows:

$$\dot{\nu} = -K_7 \nu - \bar{e}_p \tag{9}$$

with K_7 a diagonal gain matrix. The variable ν copes with the lack of measurements of $\dot{\bar{e}}$. The control design is presented through the following theorem.

Theorem 1 *Assume that the target is not vertical and the camera frame is identical with $\mathcal{R}$ (as shown on Fig.* (1)*). Let*

$$\begin{cases} T &= m\left(g + k_1 \bar{e}_3 + k_2 \nu_3\right) \\ \Gamma &= -JK_3\left(\omega - \omega^d\right) \end{cases} \tag{10}$$

with

$$\begin{cases} \omega^d &= -\frac{K_4}{g}\left(g\bar{e}_\Theta + b_3 \times \gamma^d\right) \\ \gamma^d &= -K_5\bar{e}_p - K_6\nu \end{cases} \quad (11)$$

Then,

1. *Given any upper-bound $c_M^* > 0$, there exist diagonal gain matrices $K_i = \mathrm{Diag}(k_i^j)$ $i = 3,\ldots,7; j = 1,2,3$ and scalar gains k_1, k_2, such that the control law (10) makes the equilibrium $(p, R, v, \omega, \nu) = (0, I_3, 0, 0, 0)$ of the closed-loop System (1)-(9)-(10)-(11) locally exponentially stable for any value of $c^* \in (0, c_M^*]$.*

2. *If the diagonal gain matrices K_i and scalar gains k_1, k_2 make the closed-loop system locally exponentially stable for $c^* = c_M^*$, then local exponential stability is guaranteed for any value of $c^* \in (0, c_M^*]$.* $\triangle$

This result calls for several remarks.

1) The control calculation only requires the knowledge of $\bar{H}$ (via $\bar{e}$) and ω. Thus, it can be implemented with a very minimal sensor suite consisting of a mono-camera and gyrometers only.

2) This result does not address the case of a vertical target. This case can be addressed as well with the same kind of technique and stability result. Such an extension can be found in [7] together with several other generalizations of Theorem 1.

3) Since $c^* = 1/\|X^*\|$ and $\|X^*\| \geq d^*$, a sufficient condition for $c^* \in (0, c_M^*]$ is that $d^* \geq 1/c_M^*$. Thus, Property 1) ensures that stabilizing control gains can be found given any lower bound on the distance between the reference pose and the observed planar target. This is a very weak requirement from an application point of view. Property 2) is also a very strong result since it implies that in order to find stabilizing control gains for any $c^* \in (0, c_M^*]$, it is sufficient to find stabilizing control gains for $c^* = c_M^*$. This is a much easier task which can be achieved with classical linear control tools. In particular, by using the Routh-Hurwitz criterion, explicit stability conditions on the control gains can be derived (see [7] for more details).

3.2 Nonlinear control laws

Theorem 1 shows that homography-based stabilizing control laws can be designed from very limited a priori information (essentially, a lower bound on the distance to the scene at the desired configuration and the scene planarity property). A weakness of this stability result, however, is the lack of knowledge on the size of the stability domain. Under some assumptions on the scene orientation, it is possible to derive stabilizing control laws with explicit (and large) stability domains. A first case of interest in practice is when the target is horizontal. In this case, the normal vector to the scene is known and the extraction of the orientation and position up to as scale factor, from $\bar{H}$, allows one to use available nonlinear control laws with large stability domains. Another interesting scenario for applications is when the target is vertical. This case is more challenging since knowing that the scene is vertical does not completely specify its orientation. We present below a nonlinear feedback control to address this case.

First, let us remark that $n_3^* = 0$ when the scene is vertical. Indeed, the normal vector to the scene is horizontal and the reference frame $\mathcal{R}^*$ is associated with an equilibrium configuration so that its third basis vector is vertical (pointing downward). Then, it follows from (3) that

$$\begin{cases} \sigma &:= \bar{H}b_2 \times \bar{H}b_3 - \bar{H}b_1 = R^T M(\frac{n^*}{d^*})p \\ \gamma &:= g\bar{H}b_3 = gR^T b_3 \end{cases} \quad (12)$$

with $M(\tau) = \tau_1 I_3 + S(\tau_2 b_3)$. These relations show that one can extract from $\bar{H}$ decoupled information in term of position and orientation. Compared to the result given in Proposition 1, this result is stronger since the decoupling is complete and it holds without any approximation. On the other hand, it is limited to a vertical scene. Note that γ corresponds to the components of the gravity vector in body frame. This vector, which is used in conventional control schemes based on cartesian measurements, is typically estimated from accelerometers and gyrometers measurements of an IMU, assuming small accelerations of the UAV [17].

Eq. (12) leads us to address the asymptotic stabilization of UAVs from pose measurements of the

form $\sigma = R^T M p$, $\gamma = gR^T b_3$ where M is an unknown positive definite matrix. We further assume that the velocity measurements ω and $v = R^T \dot{p}$ are also available. The variable v can be estimated, e.g., via optical flow algorithms [10, 11, 9]. In most studies on feedback control of underactuated UAVs, it is assumed that M is the identity matrix, so that the relation between the measurement function and the cartesian coordinates is perfectly known. Several control design methods ensuring semi-global stability of the origin of System (1) have been proposed in this case (see, e.g., [19, 13]). We show below that similar stability properties can be guaranteed in the case of uncertainties on the matrix M. To this end, let us introduce some notation.

For any square matrix M, $M_s := \frac{M+M^T}{2}$ and $M_a := \frac{M-M^T}{2}$ respectively denote the symmetric and antisymmetric part of M. Given a smooth function f defined on an open set of $\mathbb{R}$, its derivative is denoted as f'. Given $\delta = [\delta_m; \delta_M]$ with $0 < \delta_m < \delta_M$, we introduce the *saturating function*

$$sat_\delta(\tau) = \begin{cases} 1 \; if \; \tau \leq \delta_m^2 \\ \frac{\delta_M}{\sqrt{\tau}} - \frac{(\delta_M - \delta_m)^2}{\sqrt{\tau}\left(\sqrt{\tau} + \delta_M - 2\delta_m\right)} \; if \; \tau > \delta_m^2 \end{cases} \tag{13}$$

Note that $\tau \longmapsto \tau sat_\delta(\tau^2)$ defines a classical *saturation function*, in the sense that it is the identity function on $[0, \delta_m]$ and it is upper-bounded by δ_M.

We can now state the main result of this section (See [5] for more details, generalizations, and proof). By a standard time separation argument commonly used for VTOL UAVs, we assume that the orientation control variable is the angular velocity ω instead of the torque Γ (i.e., once a desired angular velocity ω^d has been defined, a torque control input Γ that ensures convergence of ω to ω_d is typically computed through a high gain controller).

Theorem 2 *Let sat_δ and $sat_{\bar{\delta}}$ denote two saturating functions. Assume that M is positive definite and consider any gain values $k_1, k_2 > 0$ such that*

$$\begin{cases} k_2^2 \lambda_{min}(M_s) & > & k_1 ||M_a|| ||M|| C \\ C & \triangleq & sup_\tau \left(sat_\delta(\tau) + 2\tau |sat'_\delta(\tau)| \right) \\ k_2 \delta_m & > & k_1 \\ k_1 + k_2 \delta_M & < & g \end{cases} \tag{14}$$

Define a dynamic augmentation:

$$\dot{\nu} = \nu \times \omega - k_3(\nu - \sigma), \quad k_3 > 0 \tag{15}$$

together with the control (T, ω) such that:

$$\begin{cases} \omega_1 & = & -\frac{k_4 |\bar{\mu}| \bar{\mu}_2}{(|\bar{\mu}| + \bar{\mu}_3)^2} - \frac{1}{|\bar{\mu}|^2} \bar{\mu}^T S(b_1) R^T \dot{\mu} \\ \omega_2 & = & \frac{k_4 |\bar{\mu}| \bar{\mu}_1}{(|\bar{\mu}| + \bar{\mu}_3)^2} - \frac{1}{|\bar{\mu}|^2} \bar{\mu}^T S(b_2) R^T \dot{\mu} \\ T & = & m \bar{\mu}_3 \end{cases} \tag{16}$$

where $\bar{\mu}, \mu$, and the feedforward term $R^T \dot{\mu}$ are given by

$$\begin{array}{l} \bar{\mu} := \gamma + k_1 sat_\delta\left(|\nu|^2\right) \nu + k_2 sat_{\bar{\delta}}\left(|v|^2\right) v \\ \mu := R\bar{\mu} \\ R^T \dot{\mu} = -k_1 k_3 \left[sat_\delta(|\nu|^2) I_3 + 2 sat'_\delta(|\nu|^2) \nu\nu^T \right] (\nu - \sigma) \\ \quad + k_2 \left[sat_{\bar{\delta}}(|v|^2) I_3 + 2 sat'_{\bar{\delta}}(|v|^2) vv^T \right] (\gamma - u b_3) \end{array}$$

Then,

i) there exists $k_{3,m} > 0$ such that, for any $k_3 > k_{3,m}$, the equilibrium $(\nu, p, \dot{p}, \gamma) = (0, 0, 0, g b_3)$ of the closed-loop system (1)-(15)-(16) is asymptotically stable and locally exponentially stable with convergence domain given by $\{(\nu, p, \dot{p}, \gamma)(0) : \bar{\mu}(0) \neq -|\bar{\mu}(0)| b_3\}$.

ii) if M_s and M_a commute, the same conclusion holds with the first inequality in (14) replaced by:

$$\begin{array}{rcl} k_2^2 \lambda_{min}(M_s) & > & k_1 \|M_a\| \left(\|M_a\| \sup_\tau sat_\delta(\tau) + \right. \\ & & \left. \|M_s\| \sup_\tau 2\tau |sat'_\delta(\tau)| \right) \end{array} \tag{17}$$

Let us comment on the above result. It follows from (14) that

$$|k_1 sat_\delta\left(|\nu|^2\right) \nu + k_2 sat_{\bar{\delta}}\left(|v|^2\right) v| \leq k_1 + k_2 \delta_M < g = |\gamma|$$

This guarantees that $\bar{\mu}(0) \neq -|\bar{\mu}(0)| b_3$ whenever $g b_3^T R(0) b_3 > -(k_1 + k_2 \delta_M)$. As a consequence, the

only limitation on the convergence domain concerns the initial orientation error and there is no limitation on the initial position/velocity errors. Note also that the limitation on the initial orientation error is not very strong. Note that ω_3, which controls the yaw dynamics, is not involved in this objective. Thus, it can be freely chosen. In practice, however, some choices are better than others (see below for more details).

Application to the visual servoing problem: From (12), Theorem 2 applies directly with $M = M\left(\frac{n^*}{d^*}\right) = \frac{n_1^*}{d^*} I_3 + S(\frac{n_2^*}{d^*} b_3)$. In this case one verifies that the stability conditions (14)-(17) are equivalent to the following:

$$\begin{cases} n_1^* > 0 \\ k_1, k_2 > 0 \\ k_2 \delta_m > k_1 \\ k_1 + k_2 \delta_M < g \\ n_1^* d^* k_2^2 > k_1 |n_2^*| \left(|n_2^*| + \frac{2n_1^*}{3\sqrt{3}} \right) \end{cases} \tag{18}$$

Note that the first condition, which ensures that M is positive definite, essentialy means that the camera is "facing" the target at the reference pose. This is a very natural assumption from an application point of view. When (loose) bounds are known for d^*: $d_{\min} \leq d^* \leq d_{\max}$ and $n_1^* \geq n_{1\min}$, and recalling that $|n^*| = 1$, the last condition of equation (18) can be replaced by:

$$n_{1\min} d_{\min} k_2^2 > k_1 \left(1 + \frac{2}{3\sqrt{3}} \right) \tag{19}$$

The yaw degree of freedom is not involved in the stabilization objective. On the other hand, it matters to keep the target inside the field of view of the camera. We propose to use the following control law:

$$\omega_3 = k_5 H_{21} \tag{20}$$

Upon convergence of the position, velocity, roll and pitch angles due to the other controls, the yaw dynamics will be close to $\dot{\psi} \approx -k_5 sin\psi$, thus ensuring the convergence of ψ to zero unless ψ is initially equal to π (case contradictory with the visibility assumption). Another nice feature of this yaw control is that it vanishes when $H_{21} = 0$, i.e. when the target is seen -from yaw prospective- as it should be at the end of the control task. This means that the controller tries to reduce the yaw angle only when the position/velocity errors have been significantly reduced.

4 Homography estimation

Obtaining in real-time a good estimate of the homography matrix is a key issue for the implementation of the stabilization algorithms presented before. In this section we first briefly review existing computer vision algorithms to obtain an estimate of the homography matrix. Then, we focus on the use of inertial measurements to improve and speed-up the estimation process.

4.1 Computer vision methods

There are two main classes of vision algorithms for computing the homography matrix between two images of the same planar scene:

1. Interest points based methods
2. Intensity based methods

In the first case, the homography matrix is recovered from points correspondence between the two images in a purely geometrical way. A first step consists in the detection of interest points. These correspondences can be estimate by matching (with interest point detection and descriptor) or KLT tracking (based on intensity). From this correspondence the homography matrix is recovered with algorithms such as DLT [12], which are most of the time coupled with robust estimation techniques like RANSAC or M-estimator in order to avoid false matching. For more details on interest points based methods, the reader is also referred to [12].

In the second case, the homography matrix is estimated by trying to align two images (the reference image or "template" T and the current image I). This is done, e.g., by defining a transformation (usually called "warping") from the reference image to the current image $w_\rho : q^* \longmapsto q = w_\rho(q^*)$, where q^* denotes a pixel in the reference image, q a pixel in the current image, and ρ is a parameterization of

the homography matrix, for example a parameterization of the Lie algebra of $SL(3)$. This definition leads to an optimization problem which is solved numerically. The problem consists in minimizing w.r.t. ρ a measure of the distance between the reference image $T = \{T(q^*)\}$ and the transform of the image I by the warping: $\{I(w_\rho(q^*))\}$. The cost function of the optimization problem varies with the proposed method but most of the time it essentially boil downs to a sum over the image's pixels of the distance between the pixel intensities in the two images. Usually, the optimization process only provides the optimal solution locally, i.e. provided the distance between the two images is small enough. One way to improve the convergence of this type of method is to rely on Gaussian pyramids [4]. In this case, the template image is smoothed by a Gaussian and recursively down-sampled by a factor two to form a pyramid of images, with the template image at the bottom and the smallest image at the top. The visual method is then successively applied at each level of the pyramid, from top to bottom. Thus, large movements are kept small in pixel space and the convergence domain of the method is improved.

In this paper we focus on two estimation algorithms of this second class of methods: the ESM algorithm (Efficient Second order Minimization) [3], and the IC algorithm (Inverse Compositional) [2]. Table 5.2 summarizes the main features of both methods. The main interest of the IC method is that it allows one to make a lot of precomputation based on the reference image. Indeed, the Jacobian matrix J of the cost function is computed from the template image, i.e. it depends neither on the current image nor on the homography parameterization ρ. Thus, the inverse of $J^T J$ can also be precomputed. For each iteration, only the computation of the intensity error and matrix multiplication are needed. By contrast, the ESM is a second order method that uses both the current image gradient and template image to find the best quadratic estimation of the cost function. Therefore, each iteration of the optimization algorithm is longer than for the IC method. As a counterpart, the convergence rate of the method is faster.

4.2 IMU-aided homography estimation

Cameras and IMUs are complementary sensors. In particular, cameras frame rate is relatively low (around 30Hz) and in addition vision data processing can take a significant amount of time, especially on small UAVs with limited computation power. By contrast, IMUs provide data at high frequency and this information can be processed quickly. Since IMUs are always present on UAVs for control purposes, it is thus natural to exploit them for improving the homography estimation process. We present in this section nonlinear observers recently proposed in [16] to fuse a vision-based homography estimate with IMU data. This fusion process is made on the Special Linear Lie Group $SL(3)$ associated with the homography representation (5), i.e. $\det(H) = 1$. This allows one to exploit Lie group invariance properties in the observer design. We focus on two specific observers.

The first observer considered is based on the general form of the kinematics on $SL(3)$:

$$\dot{H} = -X\,H \tag{21}$$

with $H \in SL(3)$ and $X \in \mathfrak{sl}(3)$. The observer is given by

$$\begin{cases} \dot{\hat{H}} = -\operatorname{Ad}_{\tilde{H}} \left(\hat{X} - k_1 \mathbb{P}\left(\tilde{H}(I_3 - \tilde{H}) \right) \right) \hat{H} \\ \dot{\hat{X}} = -k_2 \mathbb{P}\left(\tilde{H}(I_3 - \tilde{H}) \right) \end{cases} \tag{22}$$

with $\hat{H} \in SL(3), X \in \mathfrak{sl}(3), \tilde{H} = \hat{H}H^{-1}$. It is shown in [16] that this observer ensures almost global asymptotic stability of $(I_3, 0)$ for the estimation error $(\tilde{H}, \tilde{X}) = (\hat{H}H^{-1}, X - \hat{X})$ (i.e., asymptotic convergence of the estimates to the original variables) provided that X is constant (see [16, Th. 3.2] for details). Although this condition is seldom satisfied in practice, this observer provides a simple solution to the problem of filtering homography measurements. Finally, note that this observer uses homography measurements only.

A second observer, which explicitly takes into account the kinematics of the camera motion, is proposed in [16]. With the notation of Section 3, recall

that the kinematics of the camera frame is given by

$$\begin{cases} \dot{R} = RS(\omega) \\ \dot{p} = Rv \end{cases} \tag{23}$$

With this notation, one can show that the group velocity X in (21) is given by

$$\begin{aligned} X &= S(\omega) + \frac{vn^T}{d} - \frac{vn^T}{3d} I_3 \\ &= S(\omega) + \eta^3 \mathbb{P}(M) \end{aligned}$$

with

$$Y = \frac{vn^T}{d^*} \tag{24}$$

The following observer of H and Y is proposed in [16]:

$$\begin{cases} \dot{\hat{H}} = -\operatorname{Ad}_{\tilde{H}}\Big(S(\omega) + \eta^3 \mathbb{P}(\hat{Y}) - k_1 \mathbb{P}\Big(\tilde{H}(I_3 - \tilde{H})\Big)\Big)\hat{H} \\ \dot{\hat{Y}} = \hat{Y} S(\omega) - k_2 \eta^3 \mathbb{P}\left(\tilde{H}(I_3 - \tilde{H})\right) \end{cases} \tag{25}$$

with $\hat{H} \in SL(3), \hat{Y} \in \mathbb{R}^{3\times 3}$ and $\tilde{H} = \hat{H}H^{-1}$.

Conditions under which the estimates $(\hat{H}, \hat{Y})$ almost globally converge to (H, Y) are given in [16, Cor. 5.5]. These conditions essentially reduce to the following: *i)* ω is persistently exciting, and *ii)* v is constant. The hypothesis of persistent excitation on the angular velocity is used to demonstrate the convergence of $\hat{Y}$ to Y. In the case of lack of persistent excitation, $\hat{Y}$ converges only to $Y + a(t)I_3$ with $a(t) \in \mathbb{R}$ but the convergence of $\hat{H}$ to H still holds. The hypothesis of v constant is a strong assumption. Asymptotic stability of the observer for v constant, however, guarantees that the observer can provide accurate estimates when v is slowly time varying with respect to the filter dynamics. This will be illustrated later in the paper and experimentally verified.

4.3 Architecture and data synchronization

Implementation of the above observers from IMU and camera data is made via a classical prediction/correction estimation scheme. Quality of this implementation requires careful handling of data acquisition and communication. Synchronization and/or timestamping of the two sensor data are instrumental in obtaining high-quality estimates. If the two sensors are synchronized, timestamping may be ignored provided that the communication delay is short enough and no data loss occurs. Discrete-time implementation of the observers can then be made with fixed sampling rate. If the sensors are not synchronized, it is necessary to timestamp the data as close to the sensor output as possible, and deal with possibly variable sampling rates.

Figure 2 gives a possible architecture of the interactions between estimator and sensors (Vision and IMU). Homography prediction obtained from IMU data is used to initialize the vision algorithm. Once a new image has been processed, the obtained vision estimate, considered as as a measure, is used to correct the filter's homography estimate. Due to the significant duration of the vision processing w.r.t. the IMU sampling rate, this usually requires to re-apply the prediction process via IMU data from the moment of the image acquisition. This leads us to maintain two states of the same estimator (See Figure 2): the real-time estimator, obtained from the last homography measure and IMU data, and a post-processed estimator which is able correct a posteriori the homography estimates from the time of the last vision data acquisition to the time this data was processed.

4.4 Experimental setup

We make use of a sensor consisting of a xSens MTiG IMU working at a frequency of 200 [Hz], and an AVT Stingray 125B camera that provides 40 images of 800×600 [pixel] resolution per second. The camera and the IMU are synchronised. The camera uses wide-angle lenses (focal 1.28 [mm]). The target is placed over a surface parallel to the ground and is printed out on a 376×282 [mm] sheet of paper to serve as a reference for the visual system. The reference image is 320×240 [pixel]. So the distance d^* can be determined as 0.527[m]. The processed video sequence presented in the accompanying video is 1321 frames long and presents high velocity motion (rotations up to 5[rad/s], translations, scaling change) and occlusions. In particular, a complete occlusion of the

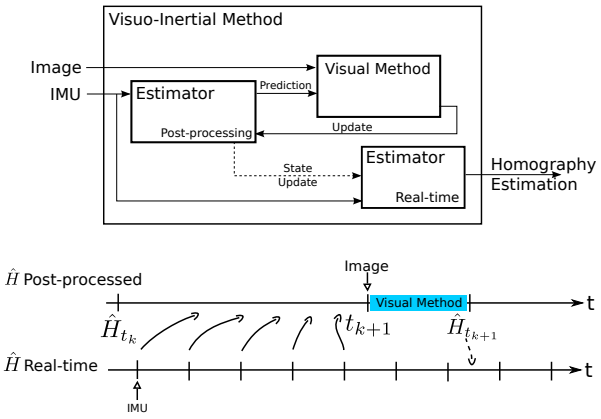

Figure 2: Visuo-Inertial method scheme and sensor measurements processing timeline

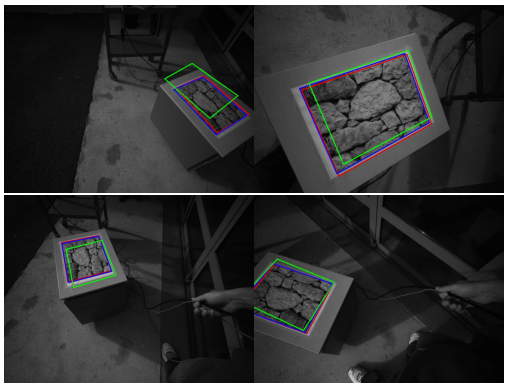

Figure 3: Four images of the sequence at 20[Hz]: pattern position at previous frame (green), vision estimate (blue), and prediction of the filterIMU (red).

pattern occurs little after $t = 10$[s].

Four images of the sequence are presented on Figure 3. A "ground truth" of the correct homography for each frame of the sequence has been computed thanks to a global estimation of the homography by SIFT followed by the ESM algorithm. If the pattern is lost, we reset the algorithm with the ground-truth homography. The sequence is used at different sampling rates to obtain more challenging sequences and evaluate the performances of the proposed filters.

For both filters (22) and (25), the estimation gains have been chosen as $k_1 = 25$ and $k_2 = 250$. Following the notation of the description available at *http://esm.gforge.inria.fr/ESM.html*, the ESM algorithm is used with the following parameter values: $prec = 2$, $iter = 50$.

4.5 Tracking quality

In this section we measure the quantitative performance of the different estimators. This performance is reflected by the number of frames for which the homography is correctly estimated. We use the correlation score computed by the visual method to discriminate between well and badly estimated frames. A first tracking quality indicator is the percentage of well estimated frames. This indicator will be labelled as "%track". Another related criteria concerns the number of time-sequences for which estimation is successful. For that, we define a track as a continuous time-sequence during which the pattern is correctly tracked. We provide the number of tracks in the sequence (label "nb track") and also the mean and the maximum of track length. Table 1 presents the obtained results for the full sequence at different sampling rates (40[Hz], 20[Hz], 10[Hz]).

The ESMonly estimator works well at 40[Hz] since 95% of the sequence is correctly tracked but performance rapidly decreases as distance between images grow (72% at 20[Hz], and only 35% at 10[Hz]). It must be noted that the ESM estimator parameters are tuned for speed and not for performance, having in mind real-time applications.

The filternoIMU estimator outperforms the ESMOnly filter on the sequence at 40[Hz]. Tracks are on average twice longer and many losses of the pattern are avoided (11 tracks versus 19 for ESMonly). At 20[Hz] the performance is still better but the difference between these two solutions reduces. At 10[Hz] the filter degrades performance.

The filterIMU tracks almost all the sequence at both 40[Hz] and 20[Hz]. There is just one tracking failure, which occurs around time $t = 10$[s] due to the occlusion of the visual target. Improvement provided by the IMU is clearly demonstrated. At 10[Hz], the performance significantly deteriorates but this filter

still outperforms the other ones.

Let us finally remark that these performances are obtained despite the fact that the assumption of constant velocity in body frame (upon which the filter stability was established) is violated.

Frame rate	Method	%track	nb track	track length	
				mean	max
40Hz	ESMonly	94.31	19	65.36	463
	FilternoIMU	97.74	11	114.27	607
1321 img	FilterIMU	98.78	2	646.5	915
20Hz	ESMonly	72.38	59	8.0	89
	FilternoIMU	80.5	52	10.17	94
660 img	FilterIMU	97.42	2	321.5	456
10Hz	ESMonly	38.79	46	2.78	27
	FilternoIMU	32.36	58	1.72	4
330 img	FilterIMU	58.66	59	3.27	27

Table 1: Rate of good track for different frame-rates and methods: percentage of well estimated frames, number of tracks, mean and maximum track length on the sequence

5 Computational aspects

Implementing vision algorithms on small UAVs is still a challenge today. Computational optimization is often necessary in order to reach real-time implementation (e.g. vision processing at about $10 - 20$ Hz). In this section, we discuss some possible approaches to speed up the vision processing for the homography estimation problem here considered.

5.1 Computational optimization

Two types of optimizations can be considered. The first one concerns the optimal use of the computing power. It consists, e.g. in computation parallelization (SIMD instructions, GPU, multiprocessor/core), fix-point computation, or cache optimization. This type of optimization does not affect the vision algorithm accuracy. Another type of optimization concerns the vision algorithm itself and the possibilities to lower its computational cost. This may affect the accuracy of the vision algorithm output. These two types of optimization have been utilized here: SIMD (Single Instruction Multiple Data) for computing power optimization, and pixels selection for vision algorithm optimization.

SIMD instructions allow one to treat data by packets. In SSE (x86 processor) and NEON (arm processor), it is possible with one instruction to treat four floating point data. So, using this instruction with careful data alignment can theoretically improve performance by a factor four. This theoretical figure is limited by load/store operation and memory (cache) transfer issues. This optimization is only done on computation intensive parts of the program such as intensity gradients computation, image warping, or Jacobian estimation.

One approach to speed up dense vision algorithms is to use only the pixels that provide effective information for the minimization process. Indeed, the lower the number of pixel, the lower the computation cost. There are many ways to select good pixels for the pixel intensity minimization between two images([8]). One approach consists in using only pixels with strong gradient since intensity errors provide position/orientation information contrary to image parts with no intensity gradient. In the experimental results reported below, we used the best 2500 pixels.

5.2 Evaluation

We report in this section experimental result obtained with both the ESM and IC methods. For each method, we uses the same stop criteria for the optimization: the maximal number of steps per scale is 30 and the stop error is 1e-3. The number of scales in the pyramid is four.

Table 2 provides the mean frame time (in ms) and mean performance (percentage of correctly estimated homographies) of the different couples of optimization and methods on the sequence at 40Hz (see experimental setup). The computation is done on a desktop PC (Intel(R) Core(TM) i7-2600K CPU @ 3.40GHz) and the same result is provided for an embedded platform (Odroid U2) based on an Exynos4412 Prime 1.7Ghz ARM Cortex-A9 Quad processor.

With SIMD the performance gain is from 3.0x to 1.7x on x86 and 1.7x to 1.17x on arm. With pixel selection the gain is better from 1.3 to 2.1 for ESM

and from 1.3x to 9x for IC.

At the end, the ratio between the fastest to the slowest is 13.6x with a lost of 22% of correctly tracked frames.

Conclusion

We have presented recent stabilization and estimation algorithms for the stabilization of VTOL UAVs based on mono-camera and IMU measurements. The main objective is to rely on a minimal sensor suite while requiring as least information on the environment as possible. Estimation algorithms have already been evaluated experimentally. The next step is to conduct full experiments on a UAV with both stabilization and estimation algorithms running on-board. This work is currently in progress. Possible extensions of the present work are multiple, like e.g. the use of accelerometers to improve the homography estimation and/or the stabilization, or the extension of this work to possibly non-planar scenes.

Acknowledgement: A. Eudes and P. Morin have been supported by "Chaire dexcellence en Robotique RTE-UPMC".

References

[1] A. Agarwal, C.V. Jawahar, and P.J. Narayanan. A survey of planar homography estimation techniques. Technical Report Technical Report IIIT/TR/2005/12, IIIT, 2005.

[2] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004.

[3] S. Benhimane and E. Malis. Homography-based 2D visual tracking and servoing. *International Journal of Robotic Research*, 26(7):661–676, 2007.

[4] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Computer VisionECCV'92*, pages 237–252. Springer, 1992.

[5] H. de Plinval, P. Morin, and P. Mouyon. Nonlinear control of underactuated vehicles with uncertain position measurements and application to visual servoing. In *American Control Conference (ACC)*, pages 3253–3259, 2012.

[6] H. de Plinval, P. Morin, P. Mouyon, and T. Hamel. Visual servoing for underactuated vtol uavs: a linear homography-based approach. In *IEEE Conference on Robotics and Automation (ICRA)*, pages 3004–3010, 2011.

[7] H. de Plinval, P. Morin, P. Mouyon, and T. Hamel. Visual servoing for underactuated vtol uavs: a linear homography-based framework. *International Journal of Robust and Nonlinear Control*, 2013.

[8] F. Dellaert and R. Collins. Fast image-based tracking by selective pixel integration. In *Proceedings of the ICCV Workshop on Frame-Rate Vision*, pages 1–22, 1999.

[9] A. Eudes, P. Morin, R. Mahony, and T. Hamel. Visuo-inertial fusion for homography-based filtering and estimation. In *IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, pages 5186–5192, 2013.

[10] V. Grabe, H.H. Bülthoff, and P. Robuffo Giordano. On-board velocity estimation and closed-loop control of a quadrotor uav based on optical-flow. In *IEEE Conf. on Robotics and Automation (ICRA)*, 2012.

[11] V. Grabe, H.H. Bülthoff, and P. Robuffo Giordano. A comparison of scale estimation schemes for a quadrotor uav based on optical-flow and imu measurements. In *IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, pages 5193–5200, 2013.

[12] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*, volume 2. Cambridge Univ Press, 2000.

		Machine	ESM		IC	
			Without SIMD	With SIMD	Without SIMD	With SIMD
Pixel Selection	No	PC	60.0(94)	20.0(94)	73.0(81)	29.5(81)
	Yes	PC	27.0(86)	15.0(86)	7.5(72)	4.4(72)
	No	odroid	347(94)	202(94)	409(81)	314(81)
	Yes	odroid	165(85)	140(86)	53(72)	45(73)

Table 2: Visual method performance: time (in ms) and accuracy (in %) for the different combination of optimization and platform.

Method	ESM	IC
Minimization objective	$\min_{\rho} \sum_{q^*} [T(q) - I(w_\rho(q^*))]^2$	
Step minimization objective	$\min_{\delta_\rho} \sum_{q^*} \left(T(q^*) - I(w_{(\rho+\delta_\rho)}(q^*)\right)^2$	$\min_{\delta_\rho} \sum_{q^*} (T(w_{\delta_\rho}(q^*)) - I(w_\rho(q^*)))^2$
Effective computation	$\delta_\rho = \left(J^T J\right)^{-1} J^T (T(q^*) - I(w_\rho(q^*)))$	
Jacobian J	$\frac{1}{2}(\Delta T + \Delta I) \left.\frac{\partial w}{\partial \rho}\right\|_\rho$	$\Delta T \left.\frac{\partial w}{\partial \rho}\right\|_0$
Use current image gradient (ΔI)	Yes	No
Use template gradient (ΔT)	Yes	Yes

Table 3: Visual method summary

[13] M.D. Hua, T. Hamel, P. Morin, and C. Samson. A control approach for thrust-propelled underactuated vehicles and its application to vtol drones. *IEEE Trans. on Automatic Control*, 54:1837–1853, 2009.

[14] F. Le Bras, T. Hamel, R. Mahony, and A. Treil. Output feedback observation and control for visual servoing of vtol uavs. *International Journal of Robust and Nonlinear Control*, 21:1–23, 2010.

[15] Y. Ma, S. Soatto, J. Kosecka, and S.S. Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. SpringerVerlag, 2003.

[16] R. Mahony, T. Hamel, P. Morin, and E. Malis. Nonlinear complementary filters on the special linear group. *International Journal of Control*, 85:1557–1573, 2012.

[17] P. Martin and E. Salaun. The true role of accelerometer feedback in quadrotor control. In *IEEE Conf. on Robotics and Automation*, pages 1623–1629, 2010.

[18] N. Metni, T. Hamel, and F. Derkx. A uav for bridges inspection: Visual servoing control law with orientation limits. In *5th Symposium on Intelligent Autonomous Vehicles (IAV 04)*, 2004.

[19] J.-M. Pflimlin, P. Souères, and T. Hamel. Position control of a ducted fan vtol uav in crosswind. 80:666–683, 2007.

[20] O. Shakernia, Y. Ma, T. Koo, and S. Sastry. Landing an unmanned air vehicle: Vision based motion estimation and nonlinear control. *Asian Journal of Control*, 1(3):128–145, 1999.

[21] D. Suter, T. Hamel, and R. Mahony. Visual servo control using homography estimation for the stabilization of an x4-flyer. 2002.