H. de Plinval
(Onera)
A. Eudes, P. Morin
(Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie)

E-mail : Henry.de_Plinval@onera.fr

DOI : 10.12762/2014.AL08-07

Control and Estimation Algorithms for the Stabilization of VTOL UAVs from Mono-Camera Measurements

This paper concerns the control of Vertical Take-Off and Landing (VTOL) Unmanned Aerial Vehicles (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 videocamera, 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 VTOL UAV feedback stabilization from homography measurements

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, especially for navigation via waypoints. Despite recent progress of this technology, especially in terms of precision, many applications cannot be addressed with the GPS as the only position sensor. First, GPS is not available indoors and it can also be masked in some outdoor environments. Then, most inspection applications require a relative localization with respect to the environment, rather than an absolute localization as provided by the GPS. Finally, evolving in dynamic environments also requires relative localization capabilities. For all of these reasons, it is important to develop control strategies based on exteroceptive sensors that can provide relative position information with respect to the local environment. 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 rich information about the environment at a relatively high frequency. 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 may 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 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 environment. 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 as a first approximation, despite the environment not being perfectly planar (e.g., as in the case of ground observation at a 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 of the displacement information between these two views [15]. This matrix can be estimated without any specific knowledge regarding the planar structure (such as 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 is the estimation of the homography matrix itself. Several algorithms have been developed within the computer vision community to obtain such an estimation (see, for example, [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 since 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 the roll and pitch control torque through fourth-order systems. For this reason, most existing control strategies based on homography measurements make additional assumptions regarding 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 § "Background". Feedback control algorithms are presented in § “Feedback Control Design” and homography estimation algorithms in § “Homography estimation”. Finally, some implementation issues are discussed in § “Computational aspects”.

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.

Control problem

Figure 1 illustrates the visual servoing problem addressed in this paper. A VTOL UAV is equipped with a mono-camera. A reference image of a planar scene $\mathcal{T}$, which was obtained with the UAV located in 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.

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 figure 1), the dynamics of these systems is described by the following well-known equations :

$$\begin{cases} m\ddot{p} = -TRb_3 + mgb_3 \\ \dot{R} = RS(\omega) \\ J\dot{\omega} = J\omega\times\omega + \Gamma \end{cases} \quad (1)$$

where p is the position vector of the vehicle’s center of mass, expressed in $\mathcal{R}^*$, $\mathcal{R}$ is the rotation matrix from $\mathcal{R}$ to $\mathcal{R}^*$, ω is the angular velocity vector of $\mathcal{R}$ with respect to $\mathcal{R}^*$ expressed in $\mathcal{R}$, $S(.)$ is 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 is the mass, T is the thrust control input, $b_3 = (0,0,1)^T$, J is the inertia matrix, Γ the torque control input and g is the gravity constant.

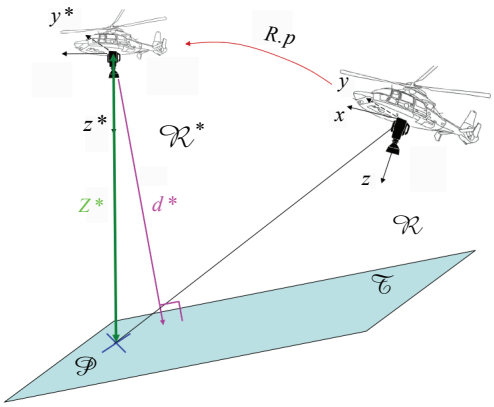

Figure 1 - Problem scheme

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 T is defined as $\{X^* \in \mathbb{R}^3 ; n^{*T} X^* = d^*\}$ where n^* are the coordinates in $\mathcal{R}^*$ of the unit vector normal to $\mathcal{T}$ and d^* is 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 . One has $X^* = RX + p$ and therefore,

$$\begin{aligned} X &= R^T X^* - R^T p \\ X &= R^T X^* - R^T p[\frac{1}{d^*} n^{*T} X^*] \\ &= (R^T - \frac{1}{d^*} R^T p n^{*T}) X^* \\ &= \bar{H} X^* \end{aligned} \quad (2)$$

where

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

The matrix $\bar{H}$ could be determined by matching 3D-coordinates in the reference and current camera planes of points of the planar scene. The cameras do not provide these 3D-coordinates, however, since only the 2D-projective coordinates of 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

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

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 the *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, for example, [15, page 135]). Therefore, α can be computed together with the matrix $\bar{H}$. Another interesting matrix is

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

Indeed, det $(H) = 1$ so that H belongs to the Special Linear Group $SL(3)$. As we will see further on, this property can be used for homography filtering and estimation purposes. Let us finally remark that

$$\eta^3 = \frac{d^*}{d}$$

Feedback Control Design

In this section, we present 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 that are affine with respect to the homography matrix components. These control laws ensure local asymptotic stabilization under very mild assumptions regarding the observed scene. The second class consists of nonlinear control laws that ensure large stability domains under stronger assumptions regarding the scene.

Linear control

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 this mixing of information must be dealt with. The control laws presented here rely on the possibility of extracting partially decoupled position and rotation information from $\bar{H}$. This is shown by the following result, first proposed in [6].

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

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

and

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

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

1. $(p,R) \to e$ *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 $(p,R) = (0, I_3)$,

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

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

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

where α^* *and* β^* *are 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 e is not complete, it is sufficient to define asymptotically stabilizing control laws, as shown below.

Let $\bar{e}_p \in \mathbb{R}^3$ (respectively $\bar{e}_\Theta \in \mathbb{R}^3$) denote the first (respectively last) three components of 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 :

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

where K_7 is 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 that the camera frame is identical to $\mathcal{R}$ *(as shown in figure 1). Let*

$$\begin{cases} T = m\left(g + k_1\bar{e}_3 + k_2\xi_3\right) \\ \Gamma = -JK_3\left(\omega - \omega^d\right) \end{cases} \tag{10}$$

with

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

Then,

1. Given any upper-bound $c_M^* > 0$, *there exist diagonal gain matrices* $K_i = Diag(k_j^i)$ $i = 3,..., 7;\ j = 1, 2, 3$ *and scalar gains* k_1, k_2, *such that the control law (10) makes the equilibrium* $(p, R, v, \omega, \xi) = (0, I_3, 0, 0, 0)$ *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 make the closed-loop system locally exponentially stable for* $c^* = c_M^*$, *then local exponential stability is guaranteed for any value of* $c^* \in (0, c_M^*)$.

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 also be addressed 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).

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 regarding the size of the stability domain. Under some assumptions regarding 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 a scale factor, from $\bar{H}$, allows available nonlinear control laws with large stability domains to be used. 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 decoupled information can be extracted from H in terms 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 the body frame. This vector, which is used in conventional control schemes based on Cartesian measurements, is typically estimated from accelerometer and gyrometer 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 $\sigma = R^T Mp, \gamma = gR^T b_3$ where M is an unknown positive definite matrix. We further assume that the velocity measurements ω and $\upsilon = R^T \dot{p}$ are also available. The variable υ can be estimated, for example, via optical flow algorithms [10, 11, 9]. In most studies on feedback control of underactuated UAVs, M is assumed to be 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, for example, [19, 13]). We show below that similar stability properties can be guaranteed in the case of uncertainties regarding the matrix M. To this end, let us introduce some notations.

For any square matrix M, $M_s = (M+M^T)/2$ and $M_a = (M-M^T)/2$ 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 with $\delta = [\delta_m, \delta_M]$ with $0 < \delta_m < \delta_M$, we introduce the saturating function

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

Note that $\tau \to \tau sat_\delta(\tau^2)$ 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

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

Define a dynamic augmentation :

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

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

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

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

$$\bar{\mu} = \gamma + k_1 sat_\delta\left(|\xi|^2\right)\xi + k_2 sat_{\bar{\delta}}\left(|v|^2\right)v$$

$$\mu = R\bar{\mu}$$

$$R^T\dot{\mu} = -k_1 k_3\left[sat_\delta(|\xi|^2)I_3 + 2sat_{\delta'}(|\xi|^2)\xi\xi^T\right](\xi - \sigma) + k_2\left[sat_{\bar{\delta}}(|v|^2)I_3 + 2sat_{\bar{\delta}'}(|v|^2)vv^T\right](\gamma - ub_3)$$

Then,

i) there exists $k_{3,m} > 0$ *such that, for any* $k_3 > k_{3m}$, *the equilibrium* $(\xi, p, \dot{p}, \gamma) = (0, 0, 0, gb_3)$ *of the closed-loop system (1)-(15)-(16) is asymptotically stable and locally exponentially stable with convergence domain given by* $\left\{(\xi, p, \dot{p}, \gamma)(0); \bar{\mu}(0) \neq -|\bar{\mu}(0)|b_3\right\}$.

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

$$k_2^2\lambda_{min}(M_s) > k_1 \| M_a \| \left(\| M_a \| sup_\tau sat_\delta(\tau) + \| M_s \| sup_\tau 2\tau | sat'_\delta(\tau) |\right) \quad (17)$$

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

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

This guarantees that $\bar{\mu}(0) \neq -|\bar{\mu}(0)|b_3$ whenever

$$gb_3^T R(0)b_3 > -(k_1 + k_2\delta_M)$$

Consequently, 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

$$M = M\left(\frac{n^*}{d^*}\right) = \frac{n_1^*}{d^*}I_3 + S\left(\frac{n_2^*}{d^*}b_3\right)$$

In this case, one verifies that the stability conditions (14)-(17) are equivalent to the following :

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

Note that the first condition, which ensures that M is positive definite, essentially 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 :

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

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 :

$$\omega_3 = k_5 H_{21} \quad (20)$$

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 is initially equal to π (case contradictory to 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 the 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.

Homography estimation

Obtaining a good estimate of the homography matrix in real-time is a key issue for the implementation of the stabilization algorithms presented earlier. 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.

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 point based methods

2. Intensity based methods

In the first case, the homography matrix is recovered from point correspondence between the two images in a purely geometrical way. A first step consists in the detection of interest points. These correspondences can be estimated by matching (with interest point detection and descriptor) or KLT tracking (based on intensity). The homography matrix is recovered from this correspondence 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 point based methods, the reader is also referred to [12].

In the second case, the homography matrix is estimated by striving to align two images (the reference image or "template" T and the current image I). This is done, for example, by defining a transformation (usually called "warping") from the reference image to the current image $w_\rho : q^* \to q = w_\rho(q^*)$, where q^* denotes a pixel in the reference image, q denotes 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 that is solved numerically. The problem consists in minimizing with respect to ρ a measurement 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 that 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 a great amount of pre-computation to be performed 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^TJ can also be pre-computed. Only the computation of the intensity error and matrix multiplication are needed for each iteration. 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.

IMU-aided homography estimation

Cameras and IMUs are complementary sensors. In particular, the camera 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 a high frequency and this information can be processed quickly. Since IMUs are always present on UAVs for control purposes, it is thus natural to make use of them to improve the homography estimation process. In this section we present nonlinear observers recently proposed in [16] to fuse a vision-based homography estimate with IMU data. This fusion process is carried out on the Special Linear Lie Group $SL(3)$ associated with the homography representation (5), i.e., $\det(H) = 1$. This allows the Lie group invariance properties to be made use of in the observer design. We focus on two specific observers.

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

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

where $H \in SL(3)$ and $X \in sl(3)$. The observer is given by

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

where $\hat{H} \in SL(3)$, $X \in sl(3)$ and $\tilde{H} = \hat{H}H^{-1}$. 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, the group velocity X in (21) can be shown to be 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 $Y = \dfrac{vn^T}{d^*}$ (24)

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

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

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

Conditions under which the estimates $(\hat{H}, \hat{Y})$ almost globally converge to (H, Y) are given in [16, Cor. 5.5]. These conditions are essentially reduced 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 $Y + a(t)I_3$ where $a(t) \in \mathbb{R}$, 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 verified experimentally.

Architecture and data synchronization

Implementation of the above observers from IMU and camera data is done via a classical prediction/ correction estimation scheme. The quality of this implementation requires careful handling of data acquisition and communication. Synchronization and/or time-stamping of the two sensor data are instrumental in obtaining high-quality estimates. If the two sensors are synchronized, time-stamping may be ignored provided that the communication delay is short enough and that no data loss occurs. Discrete-time implementation of the observers can then be done with a 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 vision estimate obtained, considered as a measurement, is used to correct the filter's homography estimate. Due to the significant duration of the vision processing with

respect to the IMU sampling rate, this usually requires the prediction process to be reapplied 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 measurement and IMU data, and a post-processed estimator that is able to correct a posteriori the homography estimates from the time of the last vision data acquisition to the time when this data was processed.

Experimental setup

We make use of a sensor consisting of an xSens MTiG IMU working at a frequency of 200 Hz and an AVT Stingray 125B camera that provides 40 images with a resolution of 800 x 600 pixels per second. The camera and the IMU are synchronized. 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 x 282 mm sheet of paper to serve as a reference for the visual system. The reference image has a resolution of 320 x 240 pixels. Thus, 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 of up to 5 rad/s, translations, scaling change) and occlusions. In particular, a complete occlusion of the pattern occurs slightly after $t = 10\ (s)$.

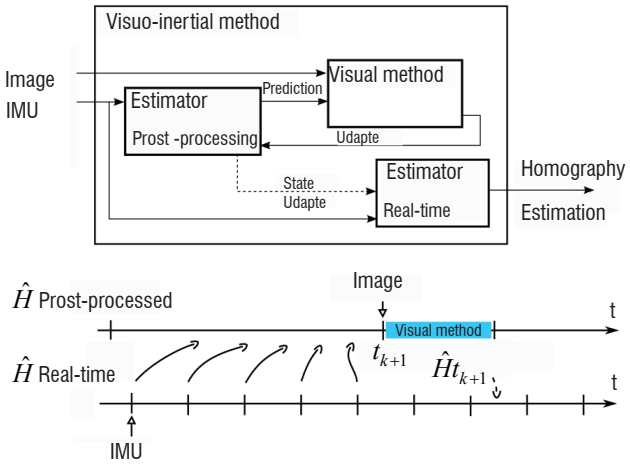

Figure 2 - Visuo-Inertial method scheme and sensor measurement processing timeline

Four images of the sequence are presented in figure 3. A "ground truth" of the correct homography for each frame of the sequence has been computed using 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 performance 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 : $prec = 2$, $iter = 50$.

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 labeled as "%track". Another related criterion concerns the number of time-sequences for which the estimation is successful. For this, 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 maximum track length. Table 1 presents the results obtained for the full sequence at various sampling rates (40 Hz, 20 Hz and 10 Hz).

The ESMonly estimator works well at 40 Hz since 95% of the sequence is correctly tracked, but performance rapidly decreases as the distance between images increases (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, with real-time applications in mind.

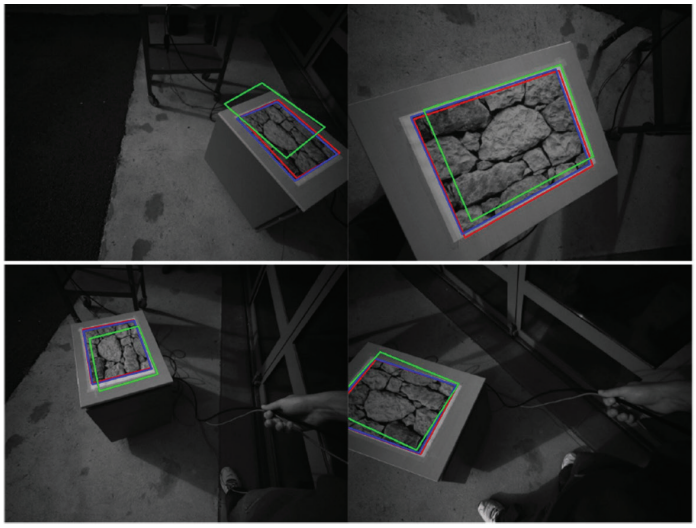

Figure 3 - Four images of the sequence at 20 Hz : pattern position at previous frame (green), vision estimate (blue) and prediction of the filterIMU (red)

The filternoIMU estimator outperforms the ESMOnly filter on the sequence at 40 Hz. Tracks are on average twice as long and many pattern losses are avoided (11 tracks versus 19 for ESMonly). At 20 Hz, the performance is even better, but the difference between these two solutions becomes smaller. At 10 Hz, the filter degrades performance.

The filterIMU tracks almost all of 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. The improvement provided by the IMU is clearly shown. 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 the body frame (upon which the filter stability was established) is violated.

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

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

Computational aspects

Implementing vision algorithms on small UAVs is still a challenge today. Computational optimization is often necessary, in order to achieve 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 considered here.

Computational optimization

Two types of optimization can be considered. The first one concerns the optimal use of the computing power. It consists, for example, 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 of lowering 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 pixel selection for vision algorithm optimization.

SIMD instructions allow the data to be processed by packets. In SSE (x86 processor) and NEON (arm processor), it is possible to process four items of floating point data with one instruction. Thus, using this instruction with careful data alignment can theoretically improve performance by a factor of 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 gradient 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 pixels, 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 a strong gradient, since intensity errors provide

		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	$\min_{\rho} \sum_{q^*} \left[T(q) - I(w_{\rho}(q^*)) \right]^2$	
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	$\delta_\rho = \left(J^T J \right)^{-1} J^T (T(q^*) - I(w_{\rho}(q^*)))$	
Jocobian J	$\frac{1}{2}(\Delta T + \Delta I) \frac{\partial w}{\partial \rho}\Big\vert_{\rho}$	$\Delta T \frac{\partial w}{\partial \rho}\Big\vert_{0}$
Use current image gradient (ΔI)	Yes	No
Use template gradient (ΔT)	Yes	Yes

Table 3 - Visual method summary

position/orientation information contrary to image parts with no intensity gradient. In the experimental results reported below, we used the best 2500 pixels.

Evaluation

In this section, we report experimental results obtained with both the ESM and IC methods. For each method, we used 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 various combinations of optimization and methods on the sequence at 40 Hz (see experimental setup). The computation is performed on a desktop PC (Intel(R) Core(TM) i7-2600K CPU @ 3.40 GHz) and the same result is provided for an embedded platform (Odroid U2) based on an Exynos4412 Prime 1.7 Ghz ARM Cortex-A9 Quad processor.

With SIMD, the performance gain is from 3.0 x to 1.7 x on x 86 and 1.7 x to 1.17 x on the arm. With pixel selection the gain is better, from 1.3 to 2.1 for ESM and from 1.3 x to 9 x for IC.

Finally, the ratio between the fastest and the slowest is 13.6 x with a loss 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 the least amount of information on the environment 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 current work are multiple, such as for example 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 d’excellence en Robotique RTE-UPMC “.

Acronyms

VTOL	(Vertical Take-Off and Landing)
UAV	(Unmanned Aerial Vehicle)
GPS	(Global Positioning System)
3D	(three-dimensionnal)
IMU	(Inertial Measurement Unit)
NED	(North-East-Down)
KLT	(Kanade-Lucas-Tomasi (feature tracker))
DLT	(Direct Linear Transformation)
RANSAC	(RANdom SAmple Consensus)
ESM	(Efficient Second-order Minimization)
IC	(Inverse Compositional)
SIFT	(Scale-Invariant Feature Transform)
SIMD	(Single Instruction Multiple Data)
GPU	(Graphics Processing Unit)

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 - *Homographybased 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*. Computer VisionECCV’92, pp. 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*. American Control Conference (ACC), pp. 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*. 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 Non-linear Control, 2013.
[8] F. DELLAERT AND R. COLLINS - *Fast Image-Based Tracking by Selective Pixel Integration*. Proceedings of the ICCV Workshop on Frame-Rate Vision, pp. 1-22, 1999.
[9] A. EUDES, P. MORIN, R. MAHONY, and T. HAMEL - *Visuo-Inertial Fusion for Homography-Based Itering and Estimation*. IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS), pages 5186-5192, 2013.
[10] V. GRABE, H.H. BÜLTHO, and P. ROBUO GIORDANO - *On-Board Velocity Estimation and Closedloop Control of a Quadrotor UAV Based on Opticalow*. IEEE Conf. on Robotics and Automation (ICRA), 2012.
[11] V. GRABE, H.H. BÜLTHO, and P. ROBUO GIORDANO - *A Comparison of Scale Estimation Schemes for a Quadrotor UAV Based on Opticalow and IMU Measurements*. IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS), pp. 5193-5200, 2013.
[12] R. HARTLEY and A. ZISSERMAN - *Multiple View Geometry in Computer Vision*. Volume 2. Cambridge Univ Press, 2000.
[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 Teedback 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 Iters 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*. IEEE Conf. on Robotics and Automation, pp. 1623-1629, 2010.
[18] N. METNI, T. HAMEL, and F. DERKX - *A UAV for Bridges Inspection : Visual Servoing Control Law with Orientation Limits*. 5th Symposium on Intelligent Autonomous Vehicles (IAV 04), 2004.
[19] J.-M. P IMLIN, P. SOUERES, 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.

AUTHORS

Henry de Plinval graduated from Ecole Polytechnique, France, in 2006 and received his MSc in Aeronautics and Astronautics in 2006 from MIT (USA), and his PhD in automatic control in 2014 from ISAE-Supaéro. Since 2008, he is with the Systems Control and Flight Dynamics Department within Onera - the French Aerospace Lab. His areas of interest include guidance, navigation and control, mostly for UAVs, with a particular focus on visual servoing for VTOL UAVs.

Alexandre Eudes received the Phd degree in robotics and computer vision from university Blaise Pascal, under the supervision of M. Lhuillier (Pascal institute) and S. Naudet (CEA). During his Phd, he worked on visual SLAM for real-time car localization applications with strong emphasis on uncertainty propagation and vision/odometry fusion. He is currently a post-doc fellow in Pascal Morin's team (ISIR), where he works on feedback control design for UAVs visual stabilization and visuo-inertial fusion for state estimation.

Pascal Morin received the Maîtrise degree from Université Paris-Dauphine in 1990, and the Diplôme d'Ingénieur and Ph.D. degrees from Ecole des Mines de Paris in 1992 and 1996 respectively. He spent one year as a post-doctoral fellow in the Control and Dynamical Systems Department at the California Institute of Technology. He was Chargé de Recherche at INRIA, France, from 1997 to 2011. He is currently in charge of the "Chaire RTE-UPMC Mini-drones autonomes" at the ISIR lab of University Pierre et Marie Curie (UPMC) in Paris. His research interests include stabilization and estimation problems for nonlinear systems, and applications to mechanical sytems such as nonholonomic vehicles or UAVs.