M. Sanfourche, J. Delaune,
G. Le Besnerais, H. de Plinval,
J. Israel, Ph. Cornic,
A. Treil, Y. Watanabe,
A. Plyer
(Onera)

E-mail: martial.sanfourche@onera.fr

Perception for UAV: Vision-Based Navigation and Environment Modeling

Nowadays, cameras and other exteroceptive sensors are on board of a large variety of automatic platforms, such as Unmanned Aerial Vehicles (UAV), space exploration probes and missiles. However, apart from this latter application, they are mostly used as payload and not to pilot the vehicle itself. In this paper, we focus on the use of computer vision for UAV perception to navigate through the environment and model it. This function is typically needed at low altitude in unknown or GPS-denied conditions. The measurements from exteroceptive sensors can then be processed to obtain information about the motion of the UAV, or the 3D structure of the environment. Our contribution is presented starting with the vision-based closed control loop, where image-based navigation is integrated to UAV control. Then, we focus on proper motion estimation techniques, like mapless relative terrain navigation or map-based GPS alternatives. Eventually, environment mapping solutions are proposed. In most cases, real image sequences coming from an aircraft or a hand-held sensor are used for validation. Our research underlines the need for new co-designed 3D sensors and massively-parallel computation technologies to go further in vision-based UAV navigation.

Introduction

Unmanned Aerial Vehicles (UAV) are mostly employed for observation or military intelligence missions. When they are not remotely operated, UAVs can only perform automatic functions such as waypoint-following, landing and take-off. For this purpose, they are equipped with high-grade inertial, GPS or radio navigation sensors. However, their navigation abilities constrain them to medium or high-altitude flight trajectories, far from any ground obstacle. The threat of an aerial collision is dealt with thanks to traffic collision avoidance systems and airspace segregation.

To extend the scope of operation of these vehicles, their safe navigation through an unknown environment or despite an intermittent or lacking GPS signal must be assured. To achieve this objective, UAVs must be equipped with exteroceptive sensors, efficient on-board computers, innovative estimation algorithms and new control laws to achieve a perception function.

In order to provide reliable information, the sensor measurements must be processed to deal with the coupling between the motion of the vehicle and the structure of the surrounding environment. Perception is also challenging because of the dependence of the measurement quality on the scene content, especially in the case of passive sensors. For instance, camera-based navigation is impossible over scenes with uniform texture, since the inference of geometrical information requires image feature association. Perception algorithms must also be robust to recurrent outlier measurements from low-level image processing, like optical flow or feature matching. Last but not least, passive sensors supply no direct information about the 3D structure of the environment, contrary to active ones like lidars, time-of-flight cameras, or Microsoft Kinect-like sensors. They must then be able to recover the 3D structure of the world with 2D-only image measurements.

Computer vision for UAVs is also challenging because of the vehicle itself. The first issue consists in the room available on board of commonly-considered platforms, such as multi-rotor or other aerial vehicles with Vertical Take-Off and Landing (VTOL) capabilities. Computers and sensors need to be placed in this very limited space, respecting weight constraints from some hundreds of grams to a few kilograms, usually no more than 2 kg [1][5][11][35][59]. Online flight computational capabilities have a direct influence on the UAV navigation performance since, contrary to a ground robot which can stop to wait computation results, a UAV is still in motion and even in hovering flight it needs to be stabilized. With its significant maneuverability and complex dynamics, it also entails a higher computational rate, which can be obtained through an Inertial Measurement Unit (IMU) aid for image processing. In practice, it is common to see solutions combining on-board computation and calculation deported to a ground

station with the main drawback of having to maintain the data link between these and the UAV [1][5]. With progress in processor architecture, however, some teams have demonstrated some computationally greedy real-time algorithms, such as Simultaneous Localization and Mapping (SLAM) or stereo-vision as in [35][59].

In practice, the misreading of the 3D environment structure is usually the most penalizing. Three strategies are possible. The easiest involves installing an active sensor on board, such as a Microsoft Kinect [46] or a flash lidar [1][5][59]. On bigger systems [19][77], lidars can be mounted on a scanning platform to offer a higher field of view. In all cases, there is a price to pay in terms of greater electric power consumption and less room available for payload instruments. The second strategy is to use a stereo rig. 3D can be inferred from it, using an algorithm running on a CPU [35] or an FPGA board [3]. The use of a Structure from Motion (SfM) algorithm with a monocular passive camera is also possible, but needs an external aid to solve for the scale [90][91]. The third and last strategy simply consists in ignoring the 3D structure of the environment. In our point of view, there is a clear separation between the techniques involved in considering or ignoring this 3D structure. The former are usually common in the navigation literature, while the latter are often seen at the control and guidance level. We must distinguish the techniques assuming a planar world, like visual servoing [20][21][70][71], from the techniques using a specific video sensor to compute the image scrolling at high rate under certain assumptions about the structure of the environment. In all cases, these techniques exploit a limited visual information, mainly the image transformation between two views, to emulate some complex behavior, such as flying down the center of a canyon [41], terrain-following [37], landing [36][81], or obstacle avoidance [11].

This article outlines recent research work at Onera regarding the perception functions for UAVs and is divided into four parts. The first one focuses on vision-based control and guidance applications. At this level, image processing and command are closely related. Image processing is designed to provide limited 2D information but at high frequency, for example, an estimation of the image motion. These techniques are illustrated through two purposes: safe landing and the rallying/stabilization of a VTOL UAV to a reference position defined by an image taken at this position.

The second part tackles the problem of UAV self-localization and 3D environment modeling relative to a local reference frame. Here, computer vision is involved in the navigation task and must infer 3D information from an image sequence. These methods are complementary to those presented in the first section. They are typically used for the flight of small UAV, indoors or outdoors, when the GPS signal is blocked or jammed. They are designed to work at video rate for several seconds, between two map position fixes. The previously described techniques are prone to trajectory estimation drift by the accumulation of small motion estimation errors.

The third part addresses computer vision techniques to provide map-based correction information similar to a GPS. The idea is to register the current sensor output with a prior map of the environment tied to a global frame. Data association between the map and the current view allows the pose of the camera to be computed and consequently the position and attitude of the vehicle. In an ideal visual navigation filter, such a function operates at a low rate, combined with a relative motion estimator working at video rate. Figure 1 illustrates our vision of a vision-aided control and navigation system and shows how the different techniques described in the article could be articulated with each other. In the last part, we focus on environment mapping techniques, which is indirectly linked to the UAV navigation.

We present an offline mapping method for environment modeling that can be used for pre-flight mission planning and for the absolute re-localization task. A second subpart is dedicated to online obstacles.

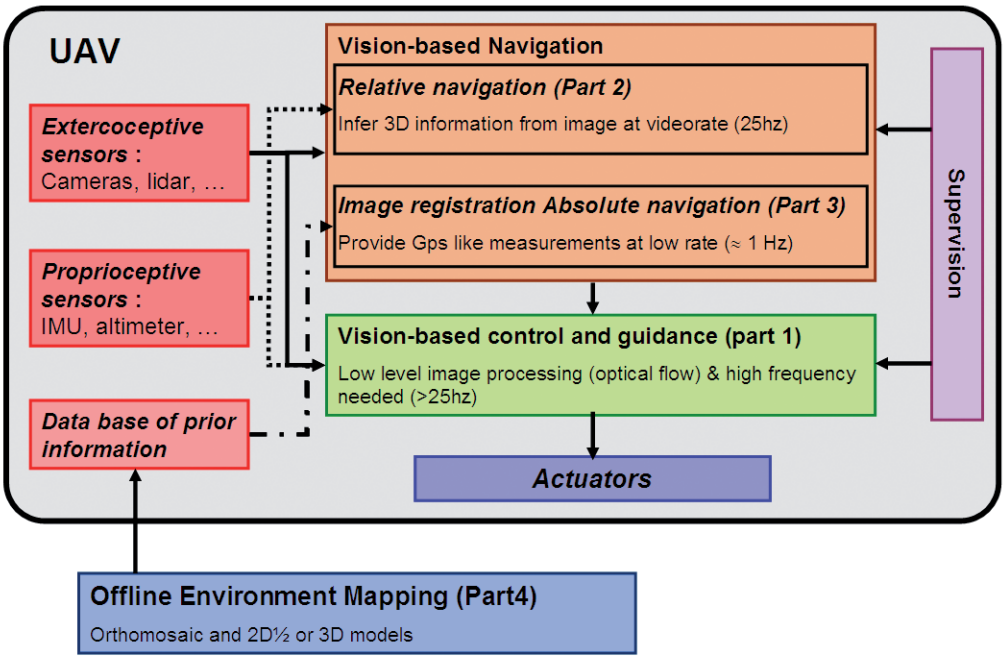


Figure 1 - Organization of a vision-based navigation and control system. Each block is related to a section of this article.

Box 1 - Optical flow

Optical flow (OF) is the field of apparent motion observed in the image plane during a video sequence. The «intensity conservation assumption», states that the visible difference between adjacent frames can be explained by «apparent motion» of pixels (or patches) from one image position to another. Actually, many image variations cannot be explained by such motion, for instance, motion blur, specular reflections, variations of illumination due to automatic gain tuning, occlusions, etc. However, one usually considers that these perturbations can either be corrected beforehand (illumination effects), or are rare (reflections, occlusions). Under the intensity conservation assumption, the optical flow derives from three contributors: the egomotion of the camera, the 3D rigid structure of the observed scene and the motion of moving objects that are in the field of view.
Figure B1-01 shows a residual optical flow norm map after global registration by a homography. Structure 3D and moving objects are very distinguishable. Optical flow then appears as a useful clue for autonomous behavior, and, indeed, it is used by most animals – including humans, of course. Its use in robotic application amounts to solving two coupled problems: (1) optical flow estimation (2) optical flow interpretation, in terms of the three components listed earlier (egomotion, 3D structure and moving parts).

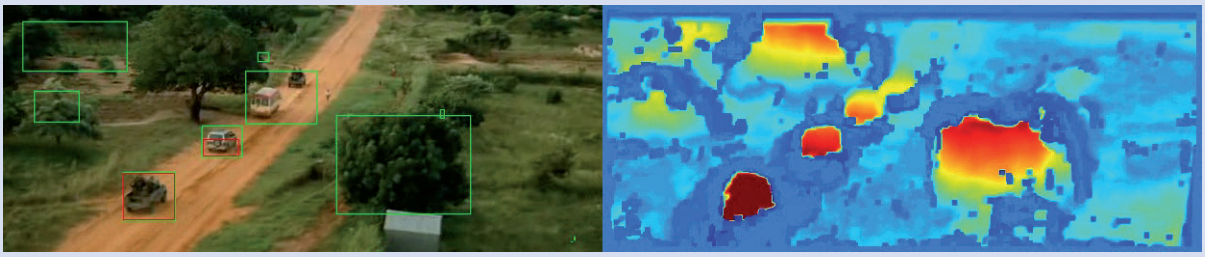


Figure B1-01 - OF norm after homographic global motion compensation. On the left, an image from the «Blood Diamond» movie. On the right, a compensated OF norm map. 3D structures and moving objects are easily detected.

Let us first comment about OF estimation. Dense OF estimation (where each pixel goes from t to $t+1?$), is an under-determinate inverse problem, which can be solved by spatial regularization. Several approaches have been proposed, ranging from costly global estimates with discontinuity-preserving properties [38][47][92] to very fast local estimates [48][54]. The latter can be obtained at video rate on full HD (2MPixels) images, thanks to FOLKIGPU [72], as illustrated in figure B1-02.

On the other hand, several applications, such as vision-based navigation, can be done with a sparse OF estimation, i.e., the estimation of the motion of a few hundred points spread over the image support – as for instance in all of the vision-based control problems of the section. Each point is chosen in a textured area, for instance using Harris-like detectors [34], and its motion is estimated by block matching techniques, often using cross-correlation maximization. These independent estimations are often improved by fitting a parametric global motion model (such as the planar homography model in the section “Online obstacle mapping for safe landing planning”) with robust techniques – in order to discard wrong matches and other outliers. This robust parametric approach is very efficient in aerial imagery, where the assumption of a planar scene is often correct above a given altitude and over a large part of the field of view.

The last issue is the interpretation of OF so as to produce quantities that are relevant for the application at hand, be it control, obstacle avoidance or target detection. Let us first consider the case of a static environment and discuss of the coupling between the scene structure and the egomotion. In the late 80s, the following first order optical development was proposed by several researchers [52][56]:

$$\begin{cases} u = f\Omega_2 - \Omega_3 y + \Omega_2 x^2/f - \Omega_1 xy/f + f T_1/Z - x T_3/Z \\ v = -f\Omega_1 + \Omega_3 x + \Omega_2 xy/f - \Omega_1 y^2/f + f T_2/Z - y T_3/Z \end{cases} \quad \text{(OF1)}$$

The left-hand sides are OF components (u,v) at pixel (x,y). In the right-hand sides, the focal length f, the translational and rotational components of the ego-motion (T and Ω) and the depth Z of the scene point whose image is projected at pixel (x,y) appears – in a referential fixed to the camera’s center. Several remarks can be made regarding these equations. First, there is a decoupling between translational and rotational effects, with the depth appearing only in the translational terms. There are two kinds of constant terms, arising either from rotation components (O1, O2) or frontoparallel translation components T_1, T_2: with a limited field of view these terms can become indistinguishable. The rotational part (three first term) is independent of the depth: it can be estimated by analytical methods, see [39][40]. Then the translational part is an affine motion model scaled by the inverse depth: the center of this model is the «focus of expansion» FOE $(fT_1/T_3, fT_2/T_3)$. An example of such a motion model can be seen in figure 21, in the case of a (mainly) translational forward motion in a canyon. The scaling by the scene structure is clearly visible in the OF norm, opening the way toward 3D reconstruction from OF estimation. Around the FOE, the OF collapses. As a general result, the use of the OF for structure and motion estimation is best done «on the sides», i.e., as far away from the FOE as the orientation of the camera and field of view allow.

Lastly, let us comment on moving object detection and characterization using the OF. In aerial imagery, as already mentioned, the scene can often be approximated by a plane and (OF1) reduced to a second order polynomial model, which can be estimated and compensated. Residual motion can then be used to detect moving objects. This is done for instance in figure B1-02. In the case of low altitude flight over a 3D area, structure effects become important and detection should integrate clues other than OF, for instance learned knowledge on the appearance of the objects that are sought.

Figure B1-02 - Optical flow on a video from the BBC movie "One day on earth". The colormap used here (source: Middlebury Optical Flow challenge [6]) at the same time codes the norm of the optical flow and the direction. This map is overlappable to the image on the right. We can easily detect the depth discontinuities.

Vision-based closed control loop

While the navigation task requires an estimation of the vehicle position and attitude, it has already been shown that coupling the output from exteroceptive sensors with a suitable control law enables automatic vehicles to achieve a complex and safe behavior, such as landing, terrain-following, flying down the center of a canyon or obstacle avoidance. We present here some work using 2D visual motion estimations and IMU measurements. First of all, we show how optical flow can be used to safely land an UAV. After that, two UAV stabilization techniques, related to visual servoing and based on the homography matrix computed between a reference view and the UAV current view, are explained.

Note, that, in both cases, a "target" must be pointed out in the image. The designation could be delivered by a human operator or by an "intelligent" embedded system.

Optical flow based control

In this section we present the control laws of a VTOL (Vertical Take-Off and Landing) UAV using optical flow as an input. A control objective is to land on a target while avoiding obstacles on its way. We first present the concept of optical flow and then describe the control laws that allow a UAV to achieve this objective.

Optical flow

Box 1 provides a detailed overview of optical flow, assuming the true image plane. For control tasks, it is more common to consider a camera as a spherical sensor, because of its passivity-like properties [32] knowing that it is possible to convert the plane image model to the spherical one [86].

Let P denote the coordinates of a point of the environment expressed in the camera-fixed frame, $\Re_D, p$ its projection on a spherical image and $\dot{p}$ its kinematics.
Then, the optical flow on a spherical image can be obtained by integrating $\dot{p}$ over a small image section W^2:

$$\phi = \iint_{W^2} \dot{p} dp$$

where W^2 is a hemisphere of the image on which we calculate the optical flow. Detailed derivation of this equation can be found in [33].

Landing on a target

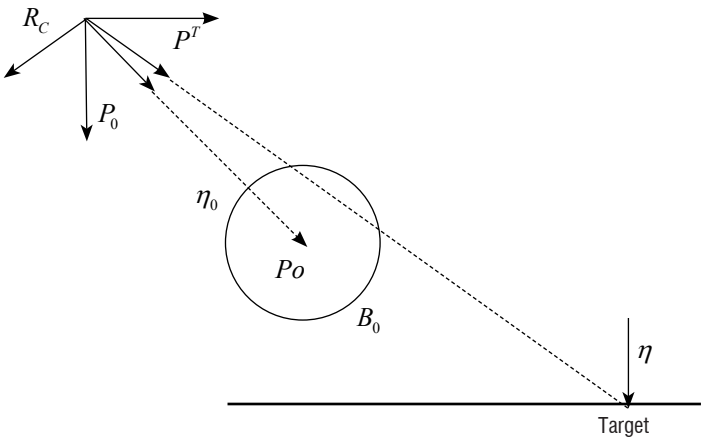


Figure 2 - Obstacle avoidance configuration

We can extract from the optical flow ϕ the translational optical flow, which is in fact the velocity relative to the ground:

$$w = -\frac{v}{d} = f(\phi)$$

where v is the UAV velocity expressed in the inertial frame. The control law for landing is designed as a PI controller, by feeding back w and the position of the target in the image. This controller allows the target to be placed in the center of the image and to descend to land.

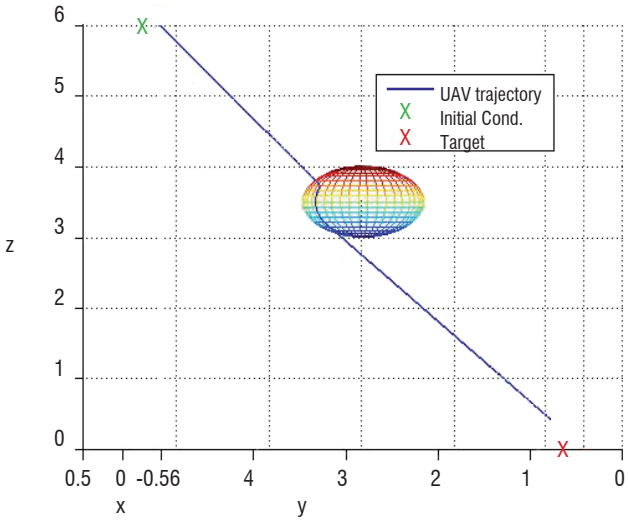


Figure 3 - Obstacle avoidance and landing. The obstacle is represented here by the repulsive sphere and we show that the UAV trajectory lays on the sphere surface

Obstacle avoidance

If the UAV crosses the repulsion sphere of an obstacle Bo during landing (figure 2), a repulsive term is activated in the control law. This repulsive part is a function of the integral of $d_o / d_o = \eta_o^t \mathrm{R}^t w_o$, which is a function of the translational optical flow of the obstacle. Choosing the repulsive gain correctly, we can guarantee that the UAV avoids the obstacle. Figure 3 shows simulation results of UAV landing trajectories, with and without obstacle.

Inner-loop vision based control for UAV stabilization

The previous parts show the benefit of an onboard video camera on the guidance/ navigation functions of a UAV. Confronted with the weaknesses of other sensors (GPS jamming, need for external devices, etc.), they also question the possibility of a UAV control system relying on a very minimal sensor suite, namely only on visual information and on other on-board sensors (typically gyrometers to measure angular velocity).

This problem is highly challenging for several reasons, some of which are related to image processing: the richness of visual information, which questions the associated computational burden, and the need for real-time computations or the required robustness to changing light conditions.

Furthermore, vision based control is also highly challenging from an automatic control prospective: depth information acts as a gain in the control loops and it cannot be extracted from images without extra knowledge; velocity is not measured; the UAV orientation is unknown unless, again, extra assumptions are made; the relationships between point coordinates seen from a camera are nonlinear, and so are helicopter dynamics, thus calling for nonlinear control techniques when addressing stability in a large flight domain. Finally, several types of UAV, such as helicopters, are not fully actuated, thus increasing the control task difficulty.

For these reasons, vision-based control for UAVs has been mostly addressed in the past years by the use of restrictive assumptions. First of all, the huge majority of works consider the observed landmarks to be lying on a plane, this assumption being necessary to use the homography matrix. Moreover, most works assume that the position and orientation (the «pose») of the UAV can be extracted from the images, leading to a more standard UAV control problem [57], or that enough knowledge is available for the dynamics to be able to be somehow inverted [31]. Several works do not consider the dynamics of the UAV (or that of another system), such as [23]. This happens for instance when the system is supposed to be fully actuated (in [13], a quadrotor is considered, where the dynamic can also be considered to be actuated).

A few recent works have addressed the nonlinearities of the UAV and camera models, often leading to local stability results [33] (the UAV has been proven to be stabilized, as long as it does not start too far off from the desired pose). Some of these works also prefer to use a spherical image camera – as mentioned in the previous part of this article dealing with optical flow – because of their passivity property, which helps in the control design steps [13].

Finally, in some recent approaches, the assumption of the knowledge of the normal to the target is made in the camera frame [50], leading to interesting but still restrictive results.

In this context, we have built on these previous works, in order to reduce the need for such assumptions still further, and have proposed two control laws:

- in the first one, a linear control law uses the homography matrix to avoid extra assumption;
- the second control law is a nonlinear control law, with an almost global stability domain.

The task at hand is to stabilize a UAV helicopter flying in front of a planar object, on which points of interest (Harris, Fast, etc.) can be extracted. The control task is to make the current image equal a reference image, supposedly taken by the UAV from the desired pose. The object nature and size, current or reference distance to object, velocity, position or orientation of the UAV are all unknown; we only have the current and reference images, as well as the gyrometer angular velocity measurements.

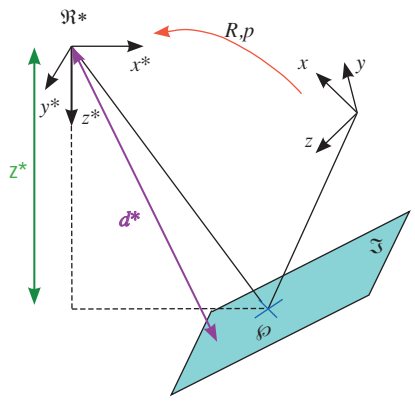


Figure 4: planar target object seen from two viewpoints

The linear control law [70]

The first important question is: how can relevant information be extracted from the images taken by the video camera? If the observed scene is a planar object seen from two viewpoints, one can estimate the so-called homography matrix, which is the transformation between the coordinates of any point of the target plane, as seen from the two viewpoints; it encompasses the rotational and translational information, while it is not possible without assumption to extract these two elements:

$$H = R^t - \frac{1}{d^*} R^t p n^{*t}$$

where R is the rotation matrix between the two viewpoints, d^* is the distance to the target plane from the reference position, p is the translation between the two viewpoints, and n^* is the normal to the target plane (see figure 4). All of these quantities are unknown in the considered scenario, so that we are left with one global measurement, which implicitly encompasses position/orientation information.
We have chosen this matrix as the core measurement for the control task: based on [9], we have defined a new error vector according to:

$$e = \begin{bmatrix} e_p \\ e_\theta \end{bmatrix} \text{ with } e_p = (I - H) m^*,\ e_\theta = vex\left(H^T - H\right) \text{ and}$$

$$vex(x) \text{ defined by } : \forall x \in \Re^3 \ : vex\left(SkewMatrix(x)\right) = x$$

$$\bar{e} = Me \text{ where } M = \begin{pmatrix} 2I & S(m^*) \\ -S(m^*) & I \end{pmatrix} \text{ and } m^* = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

In these equations, the first error vector e was defined in [9] so that it is bijectively related to the position/ orientation information. The goal of our proposed error vector $\bar{e}$ is to recover vectors that are closer to the position/ orientation information, so that a control law close to standard linear control for helicopters can be applied. In this framework, m^* is a pointing direction in the reference camera frame, which is chosen equal to the camera axis (third component of the basis). With these definitions and this choice for m^*, one shows that rendering this vector null ensures that the UAV is at the desired pose. Moreover, based on this vector, a nested loop linear control law was defined, with the addition of a dynamics augmentation. This control law is shown to stabilize the UAV without velocity measurements and without extra knowledge about the scene (size, distance, etc.) for a very wide range of values of the unknown distance to target: as long as rough bounds on this distance are known, the UAV can be stabilized. Finally, a heuristics for gain tuning was provided to fine tune the UAV performance.

The nonlinear control law [71]

Building on this first linear result, a nonlinear control law was designed. The procedure adopted was similar to that adopted in the linear case: to define an error vector recovering information close to position/orientation, then use this vector in a framework inspired by more standard helicopter UAV control. The previously-defined error vectors, although well-suited for the linear context, were not suitable for the nonlinear domain. We have defined new error vectors, once more computed from the homography matrix:

$$\pi = He_2 \wedge He_3 - He_1$$

$$\gamma = gHe_3$$

In this definition, π can be shown to be close to the position error information, whereas γ is close to the orientation error information. With these new error vectors, the error dynamics, although described with the use of unknown parameters (reference distance to target, normal vector to the target plane) and unmeasured variables (velocity), are rewritten in a form closer to a standard UAV dynamics model, which permits recent results to be used in the field of nonlinear control for helicopter UAVs [42][43]. The general form of the dynamic equations was considered, in order to prove a general result on the nonlinear robust control of an uncertain dynamic model, with the use of saturation functions. This result was applied to the vision-based control, with the aforementioned error vector, in the case of a vertical target plane. This assumption does not mean that the target plane orientation is fully known; in more recent and yet unpublished work, this assumption is forgotten. In this context, the system with such a nonlinear control law was shown to be stable for almost all initial conditions.

Future directions include the introduction of this control law into a guidance framework, which could lead to following a trajectory relying solely on video camera and gyrometer sensing.

Vision-based relative navigation

Techniques presented in the previous part exploit image measurements for guidance and control tasks. This involves a close interleaving between image processing and flight control software, because the vehicle state is not explicitly recovered by vision.

In contrast, the navigation system recovers at each time (as a minimum) the position and attitude of the vehicle. In addition, it must be able of take into account new information about the vehicle environment, for example to avoid an unknown obstacle. Here we consider the case of relative navigation, defined as navigation without a system able to locate the vehicle in an external reference frame, such as the GPS or a terrain correlator. In relative navigation, the reference frame is commonly the sensor frame at the beginning of the mission or the last available GPS statement.

The methods described here extract 3D information about the camera/ vehicle motion and the environment from a vision sensor, eventually helped by inertial sensors. These operations rely on a geometrical model of sensors, as detailed in box 2.

The first subsection describes a system combining IMU and a downward looking camera for recovering the UAV state during a target tracking scenario. In a more general manner, the two following subsections talk about visual 3D motion estimation and its accumulation over time. The estimated state provided by such techniques can be exploited to build a representation of the local environment, as described in the last subsection.

Optical flow-aided inertial navigation for ground target tracking in an urban environment

This section focuses on developing a UAV navigation and guidance system for air-to-ground target search and tracking missions in an urban environment [89]. The monocular vision-based target localization and tracking problem has been well-studied, with various applications such as aerial refueling, formation flight and ground target

BOX 2 – Camera model and calibration

One camera provides oriented information: the radiometric or/and distance information at one pixel comes from a certain direction relative to the optical axis. In order to infer geometric information from images, the process of geometrical image forming must be mathematically explained.

The direct camera model describes the image position from the 3D feature position (relative to camera position). The pinhole model, the simplest camera model, corresponds to a central projection whose center corresponds to the camera focal point. Thanks to the projective theory, this model is formulated as a linear operator, described by a 3x3 upper-triangular matrix K, where f_u, f_v, u_0, v_0 correspond respectively to the horizontal focal distance, the vertical focal distance and the coordinates of the central point projected on the image plane.

$$\mathrm{K} = \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

This model is well-suited for long-focal optics. By reducing the focal length, some geometrical distortion caused by the optics appears. To deal with it, Brown et al. [17][30] have proposed a modified model adding polynomial disturbance before the affine transform defined by the matrix K. Two sorts of distortions are identified: radial distortion, which is the majority and is described by 3 parameters, and tangential distortion, which is described by 2 parameters and describes the misalignment of the lenses w.r.t. the sensor plane. See the page "Description of the calibration parameters" on the J.Y. Bouguet Website [12].

In the case of a very large field of view, like fisheye lenses or catadioptic lenses, the Barreto [8] camera model is the best suited. This model adds 2 parameters to the previous ones.

Estimating these parameters is carried out during the calibration process. It consists of taking several images of a known geometrical pattern, like a checkerboard. Knowledge of the matching between image measurements and 3D measurements permits intrinsic parameters to be estimated thanks to a bundle adjustment process (see box 3). Some toolboxes are freely available on the Web, the most famous is due to J.Y. Bouguet [12].

For some exotic lenses, the single viewpoint condition is not respected and some authors propose to replace the parametric camera model with a look-up table pixel↔3D ray direction [73].

observation. However, most assume a UAV operation in an open space, but not in a congested area.

Two main challenges associated with an urban environment are: i) GPS signals can be degraded or even denied, and ii) there are obstacles to be avoided. Those two conditions are seldom incorporated in the UAV visual target tracking problem. Figure 5 summarizes the system that we propose to address those two issues, using onboard vision sensors. A classic monocular vision-based target localization and tracking system is augmented with optical flow-aided inertial navigation and the lidar-based obstacle mapping and avoidance algorithm.

For UAV flight safety during urban operation, it is critical to maintain its navigation capability in case of GPS signal loss. In order to limit divergence of the inertial-only navigation solution, we have suggested the use of optical flow measurement to compensate the UAV velocity information [88]. The navigation filter is based on the extended Kalman filtering method to simultaneously estimate 3D position and velocity of a target and those of the UAV. Figure 6 compares UAV trajectories estimated by the optical flow/inertial and the inertial-only navigation filters. They are calculated by using the inertial sensor measurements and the onboard camera images synchronically recorded during a UAV target tracking flight. The suggested system can provide a navigation performance equivalent to that of a GPS, while the inertial-only navigation solution diverges quickly. We are currently working on adding a ground altitude measurement with a laser for further improvement of the localization accuracy.

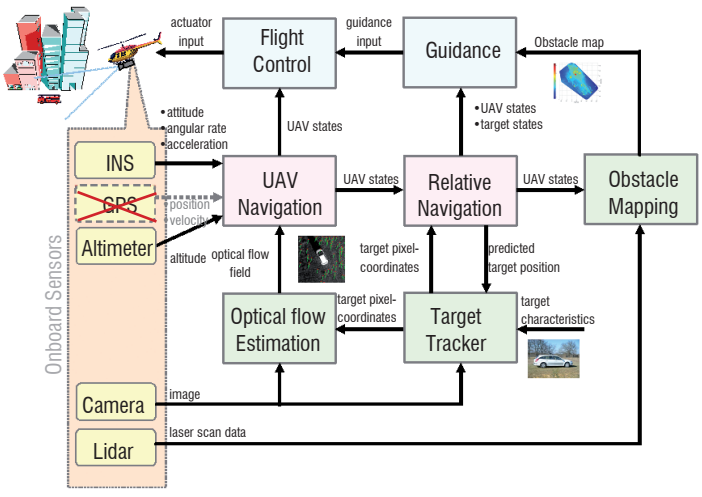


Figure 5 - Onboard navigation and guidance system for visual target tracking

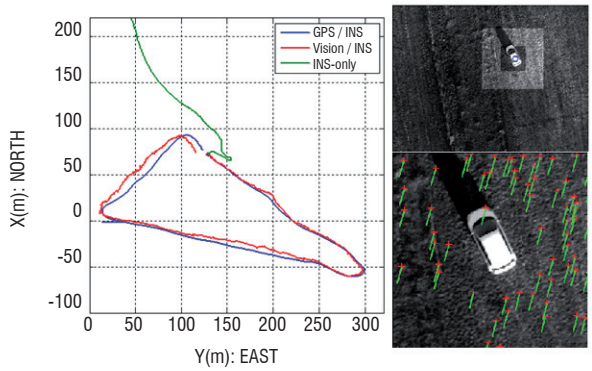


Figure 6 - Optical flow-aided simultaneous visual target tracking and navigation. The chart on the left compares three trajectories: in blue, the ground truth (from GPS/INS), in green the trajectory from inertial measurements only and in red the trajectory from our optical flow-aided navigation filter. On the right, an image with our target and optical flow.

Visual odometry

In the previous technique, the UAV state is partially recovered by vision, altitude and attitude being measured by dedicated sensors. In addition, a strong assumption was made about the environment. In contrast, visual odometry recovers the full relative motion parameters between two successive displacements in time, with the help of visual sensors. To achieve this objective, a visual odometer more commonly exploits 3D measurements from a stereorig or a depth camera. Indeed, under the commonly used hypothesis of stationary scene, the non-alignment of corresponding 3D measurements is only due to the sensor motion. Thus, recovering the sensor motion is equivalent to estimating the rigid 3D transform (rotation + translation), which aligns the two 3D measurements sets at best.

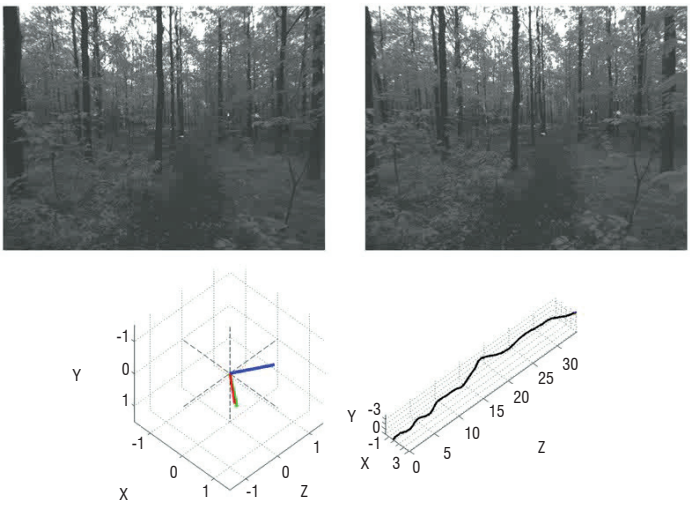


Figure 7 - Camera attitude and trajectory relative to the frame of the first image of the sequence with stereo visual odometry. On the top, one stereo image. On the bottom left, the current attitude in the reference frame; at right, the trajectory. No ground truth available, but the result is quite similar to our walk.

The most common visual odometer uses a stereorig and a sparse set of image features like Harris corners [34], FAST corners [74] or SIFT points [53]. Features, extracted from each stereo image, are matched in two ways: the stereo-matching gives the 3D localization of the feature by triangulation; the temporal matching supplies the matching between 3D measurements. From here, two strategies can be followed: the use of a 3D-3D pose computation algorithm [45] or a 2D-3D one [55]. The first minimizes a global distance expressed in the 3D Euclidean space; the second minimizes a distance on the image plane. Modern algorithms are robust to matching error through the RANSAC mechanism [29]. Figure 7 shows the result of a trajectory estimated by the real-time state-of-the-art Onera stereo visual odometer.

Stereo visual odometry can be applied on natural scenes (like in figure 7 and figure 8). In other cases – such as an indoor case, see figure 9 – a RGBD (Red Green Blue and Depth) sensor is a better choice because it can measure relative 3D information, despite the lack of texture. In [44], the authors proposed an algorithm that relies on a probabilistic framework, where a global matching criterion applied to extracted geometric features can be evaluated in parallel in the projective plane defined by the sensor camera. This algorithm benefits from a fast GPU-based implementation and has been successfully applied to Kinect data. Compared with local registration methods like the Iterative Closest Point (ICP, [10]), the proposed approach is natively robust to large camera motion (rotation or translation). The algorithm basically comprises two successive steps:

- First, sparse structures (3D contours corresponding to a list of edges) are extracted from the depth image and used in a pose score evaluation for different movement hypotheses;
- Then, a likelihood function is defined on the set of all possible transformations and used in the final decision process.

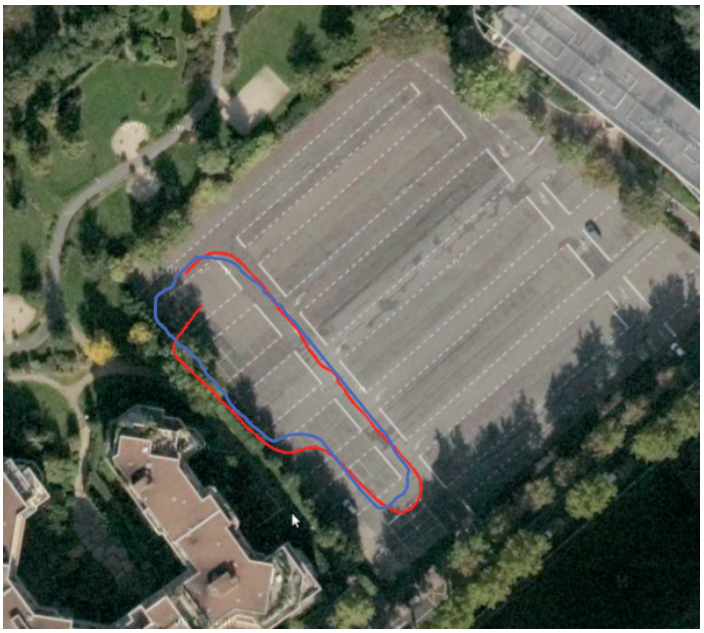


Figure 8 - Trajectory obtained by visual odometry (in red) is quite similar to the “ground truth” (blue). Note the drift: We have made a loop, but the ends of the path do not coincide.

Good results have been obtained on structured environment, as shown in figure 9. Here the transformation is selected to maximize the ML criterion. As we can see, in this case, the selected transformation permits the two point clouds to be precisely aligned. Future work will focus on coupling the pose estimation with a particle filtering approach and its embedding in a global and multi-scale SLAM framework.

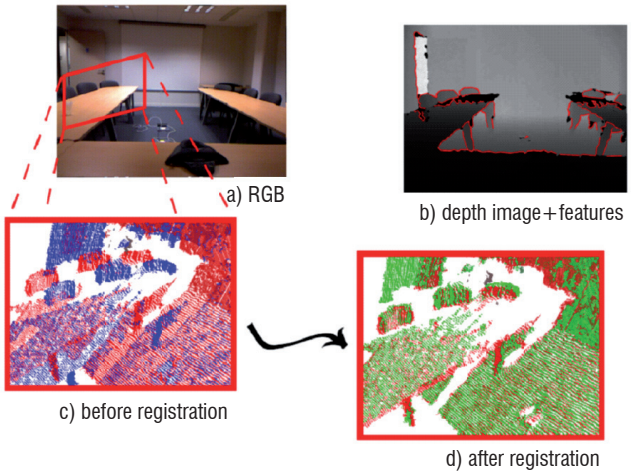


Figure 9 - Dense 3D point registration by the 3D features-based probabilistic pose developed by Onera [44]

Visual odometry is also addressed without 3D sensors like a simple monocular camera. The scale cannot be estimated directly from the sensor and external information is needed, such as an inertial measurement unit (IMU) or some known-size landmarks. We will cite Nistér et al. [64]which use the epipolar geometry and an efficient algebraic algorithm to compute the relative pose [63] and Caballero et al. [18] which use the decomposition of the homography matrix in the case of planar scenes.

vSLAM: Visual Simultaneous Localization And Mapping

A visual odometer gives the camera relative motion between two successive instants as output. The naive global trajectory estimation approach combines successive rigid transforms, accumulating unbounded estimation errors. The vSLAM algorithms attempt to reduce the drift by taking into account the geometric constraints between the sensor trajectory and some 3D landmarks tracked over time and forming a map – the Mapping term in the vSLAM acronym.

The main difficulty is due to the online nature of the algorithm: landmark positions and UAV trajectory must be refined continuously before the acquisition of the entire information. Ideally, a vSLAM algorithm must refine these parameters at each time from all of the previous algorithms, leading to an intractable formulation. In order to approximate this ideal solution, filters are designed to work on a short temporal window (from one to ten views). A great variety of filters are proposed in the SLAM literature: local bundle adjustment (see box 3) [62][78][79], Non-linear Kalman Filters (Extended, Unscented) [22] [24],or particle filter [60]. These filtering solutions have then been declined with different sensors (mono, stereo, RGBD) and with different kinds of features (point, segment, planes) and associated image processing.

Within the scope of initial work on online environment modeling, we address the problem of monocular vSLAM. We have implemented the EKF-vSLAM described in Davison [24] and Civera [22] seminal works. The state vector – containing firstly current position, attitude, linear and rotational speeds of the UAV and secondly a dynamically managed 3D landmark map– and the associated covariance matrix are updated sequentially in a prediction-measure-correction scheme. The 3D landmarks are parameterized in the inverse-depth manner [22] to simplify the non-delayed vSLAM process. With a front-mounted camera, image features and descriptors must be robust to scale and appearance changing. Unlike the original algorithm, our implementation uses DAISY descriptors [83] and we observed a better convergence on a UAV monocular image sequence, as depicted in figure 10.

Most of the previous techniques work at the standard video-rate (25-30 hz) and perhaps at a higher frame-rate (i.e., [76] announces 60 hz). Since EKF-based and Bundle Adjustment-based vSLAM need to invert a matrix whose size is related to the number of estimated parameters (viewing parameters and structure parameters), these performance are possible only by constraining the number of landmarks in the auto-generated map. In an EKF-vSLAM and standard matching techniques (i.e., without the DAISY descriptor), the video-rate is reachable with about 50 3D points in the map. This constraint prevents the use of vSLAM techniques in large environments. To bypass this limit, some techniques have propose the use of local maps regularly initialized and ordered in a graph of maps related by rigid geometrical transformation (translation, rotation) [66][68].

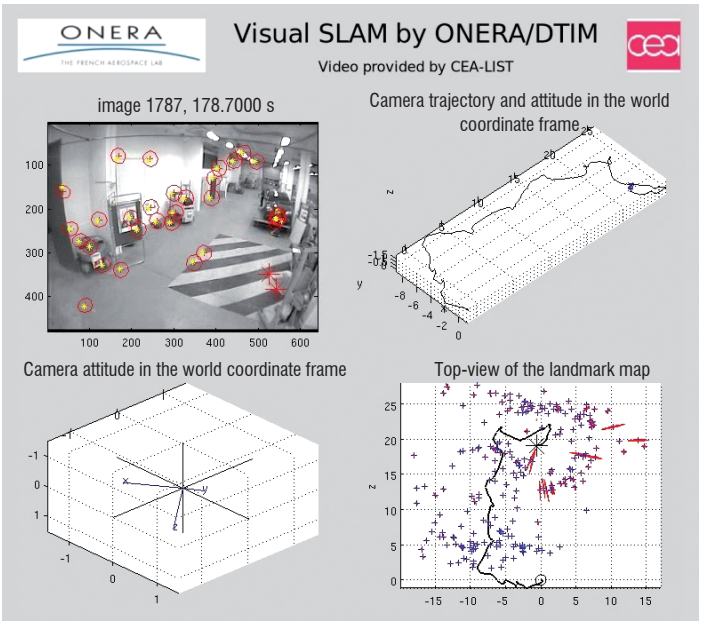


Figure 10 - DTIM EKF-vSLAM on an indoor UAV images sequence provided by the CEA-LIST

The map built by a vSLAM is not sufficient for navigation through obstacles or a congested area. In the next part, we address denser environment modeling techniques.

Online environment modeling

The data acquired by the exteroceptive sensors is also processed, in order to provide a digital model of the environment for the navigation and guidance algorithm. We describe the requirements for this model below.

Firstly, the considered application does not need a realistic and very accurate environment model – as used for view synthesis or virtual architecture. Indeed an exhaustive ternary classification of the 3D space as a free space (safe to navigate), obstacle area and unknown area is generally sufficient for planning a collision-free path. The second aspect concerns the modeling strategy. As UAV acquires

progressively more knowledge about the environment, the function must be able to combine these fragments in a sequential way, as in the SLAM manner. Lastly, the model must offer fast access to the 3D information and must be as compact as possible, in order to deal with the constraints from the embedded system (memory and computational resources are limited) .

Three environment representations could be considered: i) 3D point cloud ii) 3D meshes iii) voxel-based space quantization. Despite the unbound memory footprint, voxels-based solutions are commonly used because they natively offer fast access to information content (by three indices) and a useful neighborhood relationship between these elements. The common solution consists in computing occupancy likelihood in a stochastic update way within a voxel-based technique [2][3][27][61][93].

Two interesting references deal with the main drawback of the voxel representation. In [3], the authors propose to combine a locally classical voxel array with a global simplification stage, consisting in occupied space by polygonal convex hull. [93] shows how an octree-based voxel representation could improve the occupied memory space, without prior knowledge about the global explored volume.

In recent work, we have evaluated the Octomap solution [93] with RGBD and stereo-rig data. Figure 11 shows the 3D model obtained from a depth map provided by a home-made stereovision algorithm. Typically, the Octomap processes our data (640x480 depth maps) at 1 to 2 Hz. 90% of the computation time consists in the ray-tracing operations used for updating occupancy likelihood in the voxel grid. We have integrated a pre-processing of the depth maps proposed by [1]. The idea is to identify in the depth maps pixels corresponding to 3D points belonging to the same voxel and to replace multiple ray traces by one only. Practically, we use multi-resolution depth maps. This technique is called the “pyramidal approach”. Figure 12 shows some performance analysis results. The graph at the top compares the number of ray-tracings for different voxel resolutions with the pyramidal approach for a depth map sequence. The greater the voxel resolution, the greater the gain is. In this example, the gain in the coarsest resolution is of around100 w.r.t for the case without pyramidal approach. This result was expected. In order to more precisely qualify the impact factors, we put a lot of boxes in an office room and acquired depth maps during the tidying up (see the photo at the bottom left corner in figure 12), boxes being removed from the foreground to the background. For a same voxel resolution, the curve shows that gain also depends on the obstacles ↔ sensor distance.

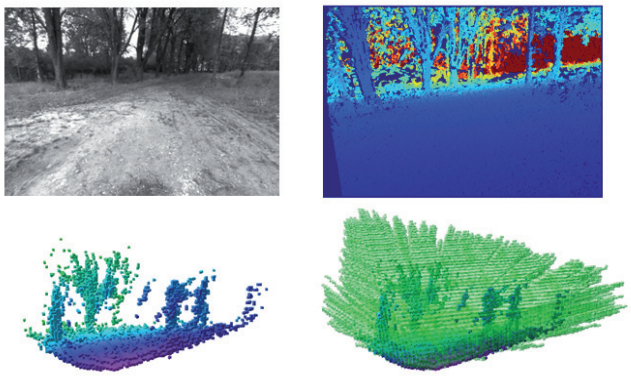


Figure 11 - Online environment modeling. From top to bottom, from left to right: grayscale stereorig image, depth map by stereovision, occupied voxels, free voxels in semi-transparent green. The obstacles within a radius of 10 meters are correctly modeled.

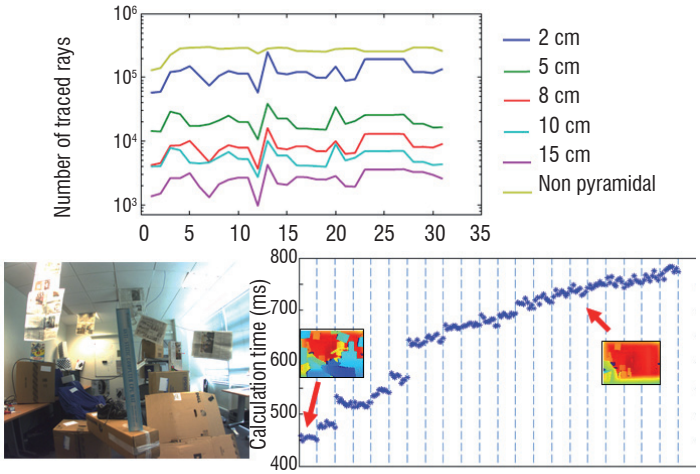


Figure 12 - The gain obtained by pyramidal approach varies with voxel resolution (at top, in terms of traced rays) and the cluttering of the scene

Image registration for absolute navigation

Vision-based relative navigation techniques presented earlier analyze the image motion of feature points to estimate the current pose of the vehicle. Except when these features stay or come back into the field of view (within a vSLAM with loop-closure detection, for example), position and attitude estimation suffers from a drift due to an unbounded error accumulation, as illustrated in figure 8. This is an issue encountered, for example, by a UAV cruising outdoors, or by a spacecraft landing on another planet from an orbit trajectory. Here, we focus on two image processing techniques to recover the vehicle position and attitude relative to a global frame. They both rely on prior geo-referenced data. The first one is a dense technique: all of the image pixels are used. The second one is sparse and relies on image feature matching.

Registration of video to geo-referenced imagery

Accurate geo-registration of video captured from an airborne platform, such as a UAV, is required for image analysis. Most of the time, the recorded data (position, attitude, zoom) supplied with the video is not sufficiently accurate for military applications. To meet this requirement, we propose an approach to automatically register airborne video to geo-referenced imagery and digital elevation models. A few manually selected “key frames” are automatically registered to geo-referenced imagery, using Mutual Information optimization [87]. This step provides ground control points that will be used to feed the “Bundle Adjustment“ procedure. The “Bundle Adjustment” (see box 3) estimates the viewing conditions (extrinsic and intrinsic parameters) of each image, using tracks of salient features generated over the sequence by the KLT tracker [80]. When the registration of a single frame is not possible because of a too narrow field of view of the video, a mosaic is built to enlarge the field of view, in order to enable registration through mutual information optimization.

Good estimates of intrinsic parameters are required to initialize the processing chain. When this is not the case, as in the videos we used, an optional “Auto-calibration” module may help to recover sufficiently

accurate focal lengths to initialize the process. The auto-calibration procedure is based on the reference [82]. The figure 13 presents the general scheme of the proposed video geo-registration chain. Figure 14 shows a result after the automatic registration of a keyframe on the orthoimage.

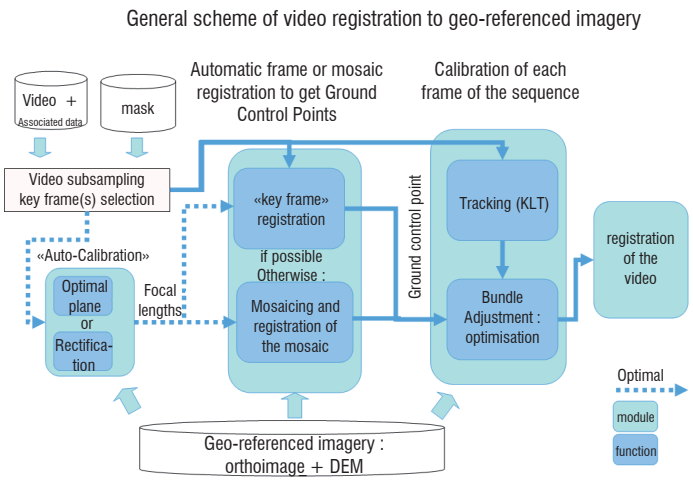


Figure 13 : Synthesis view / Flow chart of the video geo-registration method developed at Onera

A quite similar method was developed and applied to UAV navigation in [16]. In this reference, a correlation-based image registration is proposed. It differs from ours by the optimized parameters – a 2D translation versus a homography – and the maximized criterion – cross-correlation versus mutual information. Using this simplest registration model is possible, because the image registration module is combined in a complete navigation system with IMU measurements and a homography-based visual odometer, in order to adjust the UAV altitude and attitude.

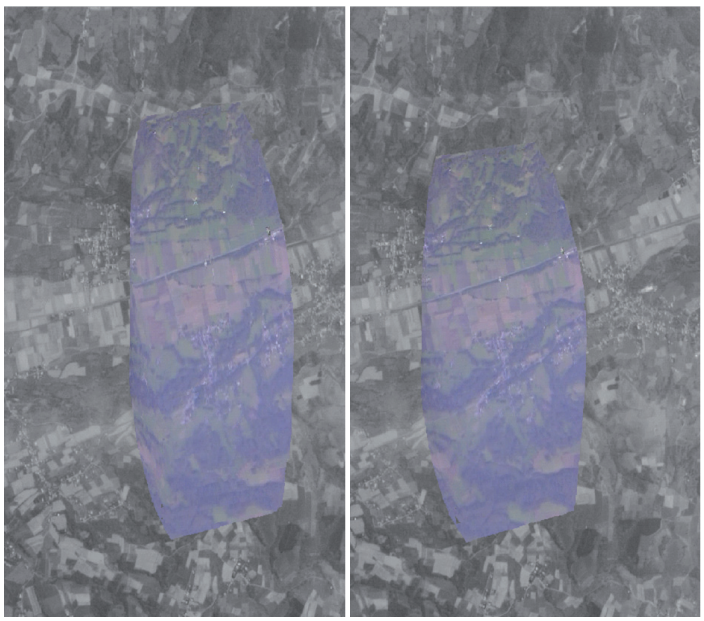


Figure 14 - Example of image registration on an orthoimage, by the alignment method based on the maximization of the mutual information. On the left, before processing. On the right, after processing, the fitting at the borders is great in most cases.

Feature-based registration for pinpoint planetary landing navigation

Pinpoint landing capability is required for several future planetary missions to the Moon and Mars. Since there are no GPS or radio beacons to provide position fixes on such planets, inertial-only systems that have been used up to date suffer from an error drift and cannot guarantee the required landing precision of 100 m. One way to reduce and maintain the navigation error low is to identify landmarks on the surface with terrain sensors.

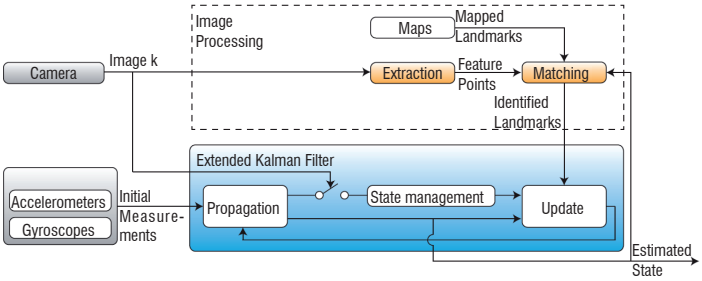


Figure 15 - Vision/Inertial fusion architecture for planetary landing

In [25], we proposed an absolute navigation algorithm that uses a simple optical camera to identify landmarks on a terrain map built from orbital data. These image-to-map matches are used as measurements in an extended Kalman filter, depicted in figure 15, which propagates inertial measurements to estimate the state in terms of position, speed, attitude and inertial biases. Inertial measurements allow high-frequency estimation to be achieved, to correct abrupt motion in the control loop. They also keep the navigation going when the camera is flown above a shadowed area, where no optical measurements are available. Vision measurements are processed at a lower rate, which is constrained by the on-board processing time of usually a few seconds. These delays are taken into account in the filter through a state management block, which adds the pose estimate of the camera at the time of image acquisition in the state of the filter and correlates it with the current estimate through the inertial measurements. Although compensated, this delay issue underlines the need for a computer-efficient visual landmark matching block, in terms of processing and memory requirements.

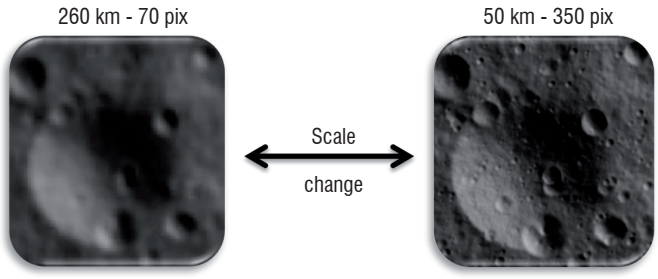
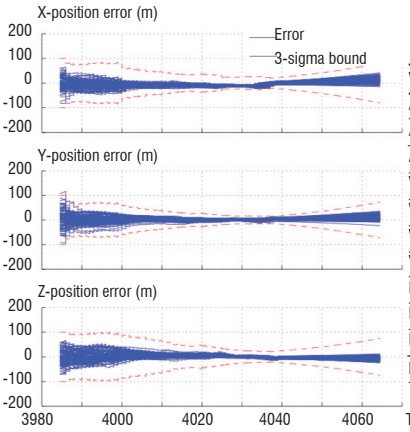


Figure 16 - Scale change illustration. The same crater is imaged by the same camera at 260 km of altitude (left) or 50 km (right)

There are three main challenges for the vision system to be used for planetary landing. The first is to be robust to illumination differences between the descent and the orbital image used to create the map. Geometric landmark descriptions, such as that proposed in [67], are more robust

than radiometric ones in this respect. They are also a lot lighter in memory requirements. The second challenge is to be able to match features on descent images shot over a broad range of altitude, with landmarks selected from orbital images taken at a constant altitude. Altitude change converts into a scale change at the image level, which is illustrated in figure 16 and is equivalent to an image smoothing. Namely, one image usually shows many more details than the other one and this is an issue to be faced by the image feature extractor. Eventually, the third vision challenge is to make use of 3D information of the world but with a 2D image only. A flat world assumption is often made in vision systems to face this problem. However, this can no longer be assumed to land on very bumpy bodies such as asteroids, or at low altitude over uneven areas on the Moon or Mars.

The vision system proposed in [67] proposes a solution to tackle these challenges. It creates a map by selecting landmarks as Harris-Laplace features on an orbital image of the area [58]. Unlike craters, these features can be found on any image of a non-uniform surface, which makes them very generic. The 3D coordinates of these landmarks are derived by back-projecting their image position in a ray that intersects a DEM of the surface. The map is an Nx5 array made up of the 3 world coordinates of each landmark, their characteristic scale on the orbital image and their cornerness scores at this scale. Online during the descent, *a priori* state estimates from the filter are used to predict which landmarks will be seen by the camera and the a priori state covariance allows an image research ellipse to be defined for each of them. Not all landmarks are selected as matching candidates, however. Only those for which the research ellipse is not overlapped by another, or those that have the highest cornerness measure of all of the overlapping ellipses, are chosen. In each pixel of the selected ellipses, the Harris cornerness measure in computed at the scale re-projected from the orbital image scale of the landmark and the position of the maximum is the descent match for the landmark. Finally, the entire set of matches is processed through the RANSAC algorithm, to look for outliers with respect to the projective camera model [29].



Variable	$3\sigma_V$	$3\sigma_{TD}$
Attitude/Yaw (deg)	0.5	0.5
Attitude/Pitch (deg)	0.7	0.7
Attitude/Roll (deg)	0.9	0.9
Velocity/X-axis (m/s)	0.7	1.2
Velocity/Y-axis (m/s)	0.8	1.2
Velocity/Z-axis (m/s)	0.9	0.9
Velocity norm (m/s)	1.0	1.1
Position/X-axis (m)	12.9	35.3
Position/Y-axis (m/s)	11.8	32.9
Position/Z-axis (m)	10.1	23.9
position norm (m)	12.0	36.1

Figure 17 - Monte-Carlo results: position error plot (left) and error statistics at the time of the last visual measurements (V) and at touchdown (TD)

This solution was tested in a 100-run Monte-Carlo analysis, dispersing the 3-sigma values of the initial errors by 1 deg, 10 m/s, and 100 m per axis respectively for attitude, speed and position. The trajectory was an Apollo-like Lunar approach, phased starting at 2000 m of altitude with a 1024x1024 camera image sensor covering a 70-deg field of view. The

BOX 3 – Bundle adjustment and aero triangulation

Bundle adjustment (BA) [84] designates the method designed to simultaneously refine the position of K 3D points and the parameters of N cameras (attitude, position and intrinsic parameters in the most general case) given image measurements. BA is formulated in a basic way as the minimization of a non-linear least squares criterion, measuring the total re-projection errors in all images:

$$J\left(\{\mathbf{C}_j\}_{j=1\ldots K},\{\mathbf{P}_i\}_{i=1\ldots N}\right)=\sum_{i}^{N}\sum_{j=1}^{K}v_{i,j}\left(\mathbf{u}_{i,j}-\mathcal{P}\left(\mathbf{C}_j,\mathbf{P}_i\right)\right)^2$$

where vi,j is a binary variable indicating if the i-th landmark is visible in the j-th view, $\mathbf{u}_{i,j}$ is the image feature corresponding to the projection of the i-th 3D landmark on the j-th image, $\mathbf{C}_j$ are the parameters of the j-th view, $\mathbf{P}_i$ are the parameters of the i-th 3D landmark and $\mathcal{P}$ is the mathematical camera model (see box 2). The maximum likelihood formulation can evidently take into account other information sources, such as prior camera positions given by (D) GPS or prior 3D landmark positions given by a map. In the same way, prior error covariance can be taken into account by replacing the L2-norm by the Mahalanobis norm. Aero-triangulation is the special case of bundle adjustment when the positions of some 3D features are known in a global frame.

The mathematical background to solve such criterion is well known (Gauss-Newton, Levenberg-Marquardt or Trust region techniques). The difficulties come from two aspects: the poor robustness of the least squares minimization scheme to outliers and the computational cost in $O(M^3)$, where M is the number of parameters to be refined simultaneously. The first problem is addressed through robust least squares, like M-estimators, which replace the L2 penalty function with an unnecessary convex function more tolerant to large residual errors. The second pitfall is by-passed by taking into account the special structure of the graph associating the refined parameters and the full set of image measurements. Indeed, since an image measurement depends only on one camera parameter and on one structure parameter, the graph is extremely sparse. This property is used to considerably reduce the computational cost. For practical purposes, the (6K+ 3N) global problem is transformed into a combination of one problem with 6K variables and N problems with 3 variables.

This algorithm is commonly called fullSLAM in the robotics community, in contrast to the EKF-SLAM which tends to solve the same criterion in a sequential way.

terrain had a 500-m height range and a 20-deg illumination difference in azimuth was introduced between the descent and orbital image, the latter being taken at a 50-km altitude. The results are shown in figure 17. All of the runs converged and touchdown 3-sigma error statistics are below 40 m in position, which is compatible with the 100-m pinpoint landing requirements. It must be noticed that below 200 m of altitude, 50 s after start, no landmark can be matched anymore, because of the limited map density and thus an inertial error drift occurs. At the end of the visual phase, the position 3-sigma error was of only 12 m, which can be considered as the actual performance of the vision system.

Environment Mapping for absolute navigation and mission-planning

Previous parts addressed vision subsystems directly involved in the control or navigation of intelligent vehicles. Mission planning and some vision-based navigation functions depend on precise maps of the flown over areas. These maps are generally provided by geographic information authorities (such as IGN in France, BkG in Germany, etc.) or by geosensing companies. However, this data could be outdated – for example, after a natural or industrial disaster –, partial– orthoimages are 2D-only, digital terrain model (DTM) do not contain objects above ground – or available in a resolution not well-suited to processing (a drawback for image registration technique for absolute navigation). To bypass these limitations, UAVs offer a low-cost solution to collect data (image, video, lidar) necessary to update maps. Note that some companies already offer this kind of service [4][69].

This section is consecrated to the environment mapping from UAVs and is illustrated through two applications. The first is relative to geosensing and photogrammetry, in order to provide DSM and orthomosaics, as in [15][26]. The second is relative to onboard mapping, to plan an unattended safe landing. In both cases, the UAVs fly above the obstacles and GPS reception is supposed to be perfect.

Offline mapping from video and lidar data

Whatever the aerial vehicle considered, an offline environment mapping task follows a well-established scheme: flight dedicated to the acquisition of heterogeneous time-stamped data (IMU,GPS, video, lidar, position of landmarks, etc.) and refinement of the trajectory estimated by the IMU/GPS navigation system thanks to exteroceptive sensors and environment modeling [15][26][49]. The proposed processing chain is no exception to the rule.

In our case, the data is acquired from the Onera UAV platform, ReSSAC [28]. This UAV is based on the Yamaha Rmax (gross mass 100 kg), equipped with a hybrid IMU/GPS-RTK navigation filter, a standard grayscale industrial camera and a 4-layer lidar scanner (SICK LD-MRS).

After a resynchronization of exteroceptive data with attitude/position information by trajectory interpolation, the video data is processed to correct the UAV 6D-trajecory (attitude and position) by bundle adjustment (see box 3 for details). Bundle adjustment uses two kinds of image features: Harris points [34] tracked by KLT [80] and SIFT [53] features matched between loop-closing frames (when the UAV fly over an area that has already been visited). The huge number of frames (more than 4000 for the example in figure 18) leads us to adopt a hierarchical bundle adjustment process. First, the trajectory is reduced to a graph of key-frames selected automatically according to an overlapping ratio deduced from the initial viewing parameters. Each node represents a key-frame and is linked to the previous and next key-frame. The graph contains some loop-closing links between temporally-distant key-frames. The loop closing detection is also based on the predicted overlap ratio between key-frames thanks to the good confidence in the UAV trajectory parameters. This graph defines the skeleton of the trajectory and a first bundle adjustment refines the parameters associated to this graph (viewing parameters of the keyframes plus position of 3D points tracked on the key-frames). In the second step, the trajectory is divided into non overlapping segments of views, delimited by key-frames and each segment is processed by a local bundle adjustment.

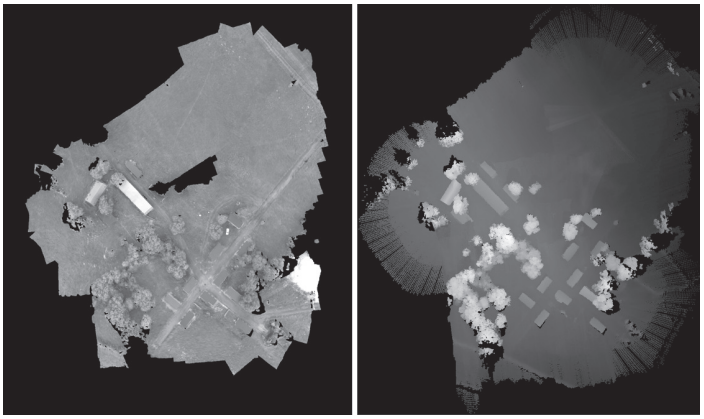


Figure 18 - Ortho-mosaic and Digital surface model of the Caylus combat camp from video and lidar data acquired by Onera ReSSAC UAV. The UAV states are refined by video aero-triangulation. Note that a lidar sensor has a greater lateral field of view than a video sensor. Thus, the ortho-mosaic does not entirely cover the DSM.

Thanks to lidar, we have access to precise 3D measurements, we do not need to infer the 3D information by image processing (higher resolution but more parameters to set). The lidar measurements are converted to 3D points localized in the global frame (relative to one point in the area selected as (0,0,0), while the axes are defined in a classical North-East-Down order). The DSM, corresponding to a regular sampling of the horizontal plane, is built by taking the altitude of the higher lidar 3D points in each cell. The orthomosaic, superimposable to the DSM, is achieved by projecting the DSM and taking into account geometric visibility, thanks to a z-buffer. Figure 18 shows the main output of our processing chain.

In its current version, our processing chain combines lidar data and video data in a suboptimal way, each sensor being used for distinguishable tasks. Work is in progress to introduce telemetric measurements within the aerotriangulation process. In the future, it would be interesting to combine these two types of data for environment modeling too.

Online obstacle mapping for safe landing planning

In the projects PRF ReSSAC [28] and PEA Action [7], a task assigned to the UAV is to explore an area to detect obstacles over ground. This obstacle map must be built online to plan a safe landing in the vicinity of an object of interest or to aid an unmanned ground vehicle in this navigation task.

For this task, we consider a monocular downward-looking camera and two restrictive assumptions: the ground is locally plane and the ground occupies quite a large part of the image. Under these assumptions,

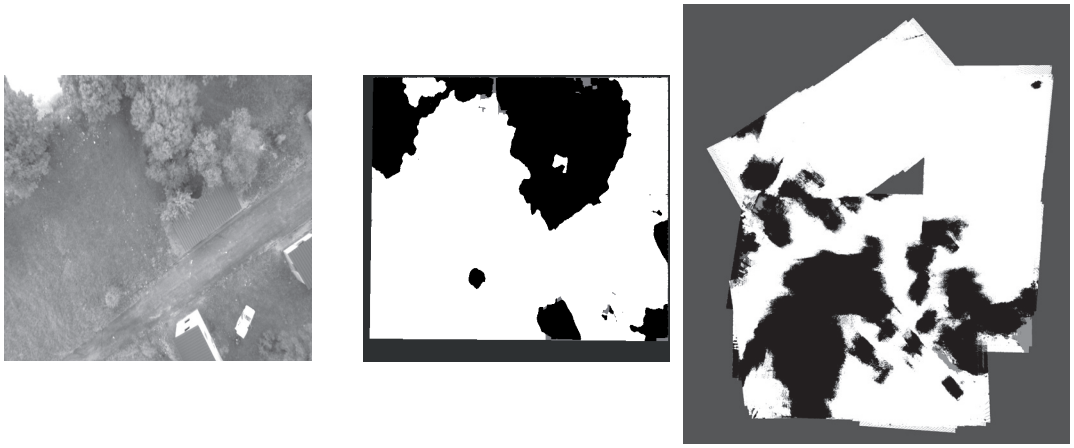


Figure 19 - Detection of obstacles above ground by segmenting optical flow. On the left: the reference image; in the center, the detection map (white: safe; black: obstacle; light gray: unknown; dark gray: non-overlapping regions)

obstacle detection is equivalent to finding an image area for which image motion (between two successive views) cannot be described by the majority homography. Our work is quite similar to [14]. As in this latter reference, a first step uses the matching of Harris corners [34] between two views to infer the majority homography, thanks to a robust least square method (Least Median of Squares[75]). Our method differs from [14] with the obstacle classification step. In [14], authors compute the correlation between the reference view and the second image warped into the reference, thanks to estimated homography (this warped image is a so-called displaced frame difference). On our side, we have developed two concurrent methods, based on the comparison of the "true" optical flow and the estimated homographic optical flow. The first is based on the Odobez work [65]: the proposed criterion computed from the DFD approximates the previous mentioned difference. The second uses the CPU-based efficient quasi-dense optical flow algorithm proposed by Lhuillier et al in [51]. In this case, the detection is directly obtained by optical flow comparison. In both cases, a threshold criterion produces ternary maps (obstacle / free / unknown). The computation times of these two methods are similar, between 2 and 3 frames per second on a 2.5Ghz Intel Core2 Duo (the algorithm is not multi-threaded) on 640x480 images. The processing may seem slow, however, this frequency permits a better baseline to be acquired between two successively processed frames. Figure 16 presents the result of our method, based on the Quasi-dense optical flow. Here, the ground occupies a large part of the image and the detection looks precise, the hole between the house and the trees (near of the image center) is detected as non-obstacle. These individual detection maps are then associated in a global map, thanks to viewing parameters to drive a motion planning algorithm (map on the right in figure 19).

The same video sequence has been processed to obtain the obstacle map on the right in figure 19 and the DSM in figure 18. A comparison could be made and the estimated obstacle map is globally consistent with the DSM – at the same time, in the ground area and in the above-ground elements – despite the enlargement of the obstacle footprint (the pixels are projected onto the map under the hypothesis of a plane scene) or some misalignment due to erroneous viewing parameters.

Conclusion

We have presented here several contributions for navigation, mainly carried out in the context of Onera's research projects PRF SPIDER and PR AZUR, more precisely described in box 4. We have shown that the three control levels of an automatic system could require perception functions. This interaction between the control and the perception functions involves the processing time and, consequently, the nature of the information provided by the image processing.

For the lowest control level, the image processing provides 2D-only information, such as optical flow (dense or parametric) or parametric image transform (homography). The emphasis is placed on control aspects. At an intermediary level, we are interested in the vision-based relative navigation. Videos are processed in order to provide a geometrically-consistent trajectory and environment modeling. This 3D information must be inferred from passive or active sensors. This topic is addressed outside the context of vehicle control. Lastly, we have presented some work for replacing GPS measurements and for long-term mapping. Replacing GPS requires reference maps to register an image on it. As is frequently the case in image processing, the problem can be addressed by using image features or the whole image. Long-term mapping provides a useful model for – online or offline – mission planning and for GPS replacement. By taking into account all available measurements in a global process, offline mapping offers the more precise results.

In the future, two technologies will make it possible to go further in the autonomous navigation of aerial vehicles.

The first concerns the generalization of massively-parallel architectures. Image processing is well suited to this kind of architecture because many of the techniques rely on image filtering, for example to identify image features. Thanks to increasingly energy-efficient electronics, solutions are being combined on the same silicon piece multi-core CPU and a Graphical Processing Unit (GPU), promising increasing computing power.

The second concerns the sensors and especially the 3D sensors. The Microsoft Kinect is a good example of a technological breakthrough. It is a lightweight and energy-efficient device that has been largely adopted for indoor robotics (aerial or not). Its main drawback is its active nature, which limits its usage domain to the indoor one. Also, for outdoor use, the development of 3D compact passive sensors, like the one proposed in the SPIDER project (see box 4), seems a very promising way ■

BOX 4 – Perception for navigation themes at Onera

The themes of perception for automatic vehicle navigation are articulated around two internal research projects with separate finalities. The PRF SPIDER (first subsection) is concerned with the development of perception components (sensor and calculators) useful for UAV navigation. The PR AZUR (second subsection) is oriented onboard integration and in-flight demonstration with specific development around control and guidance using perception algorithms.

SPIDER: co-designing the «eyes» of future micro-UAVs

A major issue in micro-UAV concepts is to design their eyes, i.e. integrated «perception devices», made of sensors and microcomputers, which should not only provide good images, but also directly provide good information for autonomous behavior. This is the goal of Onera's internal project SPIDER, a French acronym which could be translated as «Small Perception and Interpretation Devices for Urban EnviRonment». To start with, SPIDER integrates researchers in the domains of optics, vision and control and promotes a «co-design» approach. Co-design amounts to evaluating the end-to-end performance of a perception device, including optics, vision algorithm and control parameters, so as to search for a global optimum to the perception problem at hand. This approach tends to reduce the overfitting issues associated with traditional sequential design. The main objective of SPIDER is to produce demonstrations of the proposed perception device concepts. Among the most interesting SPIDER demonstrators is CAM3D, a passive monocular sensor with 3D capability, based on the «depth-from-defocus» paradigm: preliminary results, extracted from the PhD work of Pauline Trouvé (Onera) [85] are presented in figure 22. This figure shows a realistic and coherent raw depth estimation, using a criterion called GL developed in the DTIM. It shows a good localization of depth discontinuities in most cases. Even the incised trunk is detected. The median filter efficiently removes the noise, with some artifacts on the discontinuities.

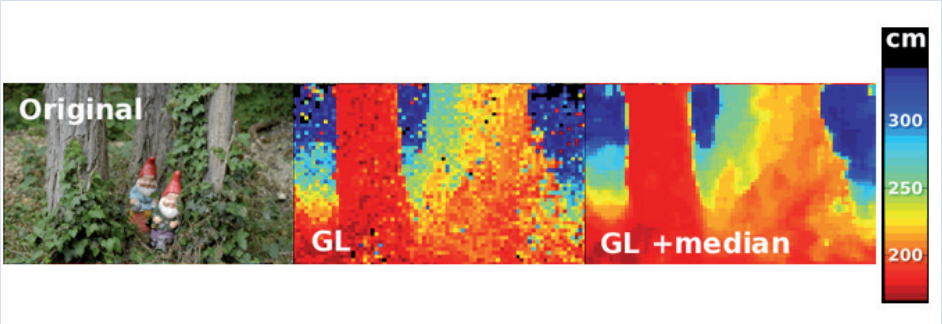


Figure B4-01 - Output of the CAM3D concept. On the left, the RGB image. In the center, the raw depth map after a depth-from-defocus process. On the right, the depth map after a median post-processing. Depth varies between 1.5m and 3.5m according to the indicated colormap

AZUR: Autonomous navigation of UAV in an urban zone.

The AZUR project is aimed at making an onboard navigation software system of a VTOL-type UAV for its fully-autonomous/semi-autonomous operation in an urban environment. As illustrated in the figure below, the system is a closed-loop chain of perception, decision and action. In order to ensure flight safety, onboard perception is mandatory to obtain the current situation of both the UAV and of the environment (obstacles, wind conditions, etc.). In the AZUR project, the following four function modules will be developed:

- Real-time environment mapping and path planning;
- Obstacle detection and reactive avoidance;
- Navigation without GPS;
- Wind gust estimation and compensation.

Especially, the first three apply active and/or passive vision-based control approaches. For example, optical flow-based visual servoing for obstacle avoidance, and visual odometry for GPS-free navigation.

These modules will be integrated into one complete navigation system and implemented onboard one or more of the Onera UAV experimental platforms. Flight demonstration of their safe autonomous operation in an obstacle field is expected at the end of the project.

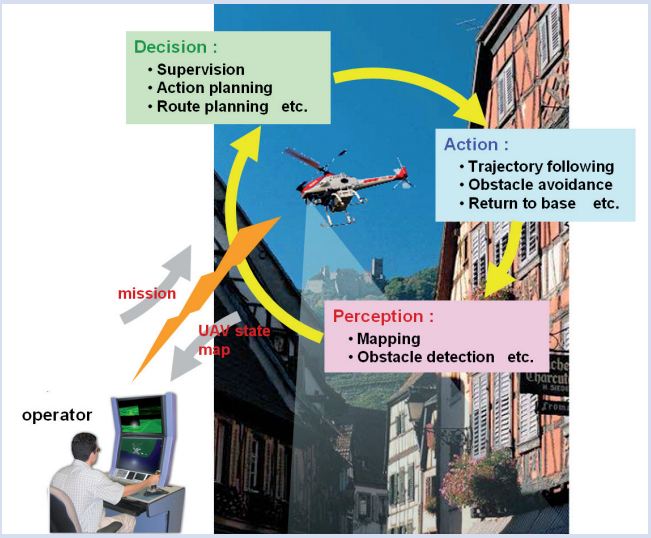


Figure B4-02 - UAV on-board navigation system

Acronyms

BA (Bundle Adjustment)
CAM3D (CApteur passif Monovoie à capacité 3D)
CPU (Central Processing Unit)
DEM (Digital Elevation Model)
DSM (Digital Surface Model)
EKF (Extended Kalman Filter)
FPGA (Field-Programmable Gate Array)
GPS (Global Positioning System)
GPU (Graphics Processing Unit)
IMU (Inertial Measurement Unit)
OF (Optical Flow)
PR AZUR (Projet de Recherche Autonomie en Zone URbaine)
PRF SPIDER (Projet de Recherche Fédérateur Système de Perception et d'Interprétation Dynamique en Environnement uRbain)
SLAM (Simultaneous Localization and Mapping)
RGBD (Red-Green-Blue-Depth (camera))
UAV (Unmanned Aerial Vehicle)
VTOL (Vertical Take-Off and Landing)

References

[1] M. ACHTELIK, A. BACHRACH, R. HE, S. PRENTICE, N. ROY - *Autonomous Navigation and Exploration of a Quadrotor Helicopter in GPS Denied Indoor Environments*. Proceedings of International Aerial Robots Competition, 2009

[2] F. ANDERT - *Drawing Stereo Disparity Images Into Occupancy Grids : Measurement Model and Fast Implementation.* Proceedings of IEEE International conference on Intelligent Robots and Systems, 2009

[3] F. ANDERT, F. ADOLF - *Online World Modeling and Path Planning for an Unmanned Helicopter.* Autonomous Robots, Vol. 27(3), pp 147-164 , 2009

[4] L'Avion Jaune - http://www.lavionjaune.fr/

[5] A. BACHRACH, S. PRENTINCE, R. HE, N. ROY - *RANGE: Robust Autonomous Navigation in GPS-Denied Environments.* Journal of Field Robotics, Vol 28(5), pp 644-666, 2011

[6] S. BAKER, D. SCHARSTEIN, J.P. LEWIS, S. ROTH, M. BLACK, R. SZELISKY - *A Database and Evaluation Methodology for Optical Flow.* International Journal on Computer Vision, Vol 92(1), pp 1-31, 2011

[7] M. BARBIER, S. LACROIX and al. - *PEA Action, Research Project Granted by French DGA*. Website : http://action.onera.fr/

[8] J.-P. Barreto - *A Unifying Geometric Representation for Central Projection Systems.* Comuter Vision and Image Understanding, Vol. 103(3), pp 207-217, 2006

[9] S. BENHIMANE, E. MALIS - *Homography-Based 2D Visual Tracking and Servoing*. International Journal of Robotics Research, Vol XX(yy), pp 661–676, 2007

[10] P. Besl, H. McKay - *A Method for Registration, of 3-D Shapes.* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, N°2, pp 239-256, 1992

[11] A. BEYELER, J.-C. ZUFFEREY, D. FLOREANO - *Vision-Based Control of Near-Obstacle Flight.* Autonomous Robots, Vol. 27 (3), pp 201–219, 2009

[12] J.Y. BOUGUET - *Camera Calibration Toolbox for Matlab*. Dowloadable on http://www.vision.caltech.edu/bouguetj/calib_doc/index.html, 2010

[13] O. BOURQUARDEZ, R. MAHONY, N. GUENARD, F. CHAUMETTE, T. HAMEL, L. ECK - *Image-Based Visual Servo Control of the Translation Kinematics of a Quadrotor Aerial Vehicle.* IEEE Transactions on Robotics, 2009

[14] S. BOSCH, S. LACROIX, F. CABALLERO - *Autonomous Detection of Safe Landing Areas for an UAV from Monocular Images.* IEEE International Conference on Intelligent Robots and Systems, 2006

[15] M. BRYSON, A. REID, F. RAMOS, S. SUKKARIEH - *Airborne Vision-Based Mapping and Classification of Large Farmland Environments.* Journal of Field Robotics, Vol. 27(5), pp. 632-655, 2010

[16] G. CONTE, P. DOHERTY - *Vision-Based Unmanned Aerial Vehicle Navigation using Geo-Referenced Information.* Eurasip Journal of advances in signal processing, 2009

[17] D.C. BROWN - *Decentering Distortion of Lenses.* Photogrammetric Engineering and Remote Sensing, Vol. 32(3), pp 444-462, 1966

[18] F. CABALLERO, L. MERINO, J. FERRUZ, A. OLLERO - *A Visual Odometer Without 3D Reconstruction for Aerial Vehicles*. Applications to building inspection, in proceedings of IEEE International Conference on Robotics and Automation, 2005

[19] L. CHAMBERLAIN, S. SCHERER, S. SINGH - *Full-Scale Automated Landing and Obstacle Avoidance in Unmapped Environments.* Proceedings of the 67th forum of the American Helicopter Society, 2011

[20] F. CHAUMETTE, S. HUTCHINSON - *Visual Servo Control Part 1: Basic Approaches.* IEEE Robotics and Automation Magazine, Vol. 13(4), pp 82-90, 2006

[21] F. CHAUMETTE, S. HUTCHINSON - *Visual Servo Control Part 2: Advanced Approaches.* IEEE Robotics and Automation Magazine, Vol. 14(1), pp 109-118, 2007

[22] J.CIVERA, A. DAVISON, J.M.M. MONTIEL - *Inverse Depth Parametrization for Monocular SLAM*. IEEE Transactions on Robotics, Vol. 24(5), pp 932-945, 2008

[23] R. CUNHA, C. SILVESTRE, J. HESPANHA, A.P. AGUIAR - *Vision-Based Control for Rigid Body Stabilization*. Automatica, 2011

[24] A. DAVISON, I. REID, N. MOLTON, O. STASSE - *MonoSLAM: Real-Time Single Camera SLAM.* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29(6), pp 1052-1067, 2007

[25] J. DELAUNE, G. LE BESNERAIS, M. SANFOURCHE, T. VOIRIN, C. BOUDARIAS, J.-L. FARGES - *Optical Terrain Navigation for Pinpoint Landing,* Proceedings of 35th AAS Annual Guidance and Control Conference 2012

[26] H. EISENBEISS - *UAV Photogrammetry, PhD Thesis of Institute of Geodesy and Photogrammetry.* ETH Zurich, 2009

[27] A. ELFES - *Using Occupancy Grids for Mobile Robot Perception and Navigation.* Computer, Vol. 22(), pp , 1989

[28] P. FABIANI, A. PIQUEREAU, V. FUERTES and al. - *Projet de Recherche Fédérateur ReSSAC*. Onera Research project. Website : http://www.onera.fr/dcsd/ressac/index.html

[29] M. A. FISCHLER, R. C. BOLLES - *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography.* Comm. of the ACM, Vol 24, pp 381-395, 1981

[30] J.G. FRYER, D.C. BROWN - *Lens Distortion for Close-Range photogrammetry.* Photogrammetric Engineering and Remote Sensing, Vol. 52(1), pp 51-58, 1986

[31] T. GONÇALVES, J.R. AZINHEIRA, P. RIVES - *Vision-Based Autonomous Approach and Landing for an Aircraft Using a Direct Visual Tracking Method*. IEEE Conference on Robotics and Automation, 2010

[32] T. HAMEL, R. MAHONY - *Visual Servoing of an Under Actuated Dynamic Rigid-Body System : an Image-Based Approach.* IEEE Transactions on Robotics and Automation, Vol. 18(2), 2002

[33] T. HAMEL, R. MAHONY - *Image Based Visual Servo-Control for a Class of Aerial Robotic Systems*. Automatica, 2007

[34] C. HARRIS, M. STEPHENS - *A Combined Corner and Edge Detector.* Proceedings of the 4th Alvey Vision Conference, 1988

[35] L. HENG, L. MEIER and al. - *Autonomous Obstacle Avoidance and Maneuvering on a Vision-Guided MAV Using on-Board Processing*. proceedings of International Conference on Robotics and Automation, 2011

[36] B. HERISSE, F. RUSSOTTO, T. HAMEL, R. MAHONY - *Hovering Flight and Vertical Landing Control of a VTOL Unmanned Aerial Vehicle Using Optical Flow.* proceedings of International conference on Intelligent Robots and Systems, 2008

[37] B. HERISSE, T. HAMEL, R. MAHONY, F.-X. RUSSOTO - *A Terrain-Following Control Approach for a VTOL Unmanned Aerial Vehicle Using Average Optical Flow.* Autonomous robots, Vol. 29(3), pp 381-399, 2011

[38] B. HORN, B. SCHUNCK - *Determining Optical Flow*. Technical Report, Massachusetts Institute of Technology, April 1980

[39] B.K.P. HORN - *Closed-Form Solution of Absolute Orientation Using Unit Quaternions*. Journal of Optical Society of America, A4, pp 629-642, 1987

[40] B.K.P. HORN, H.M. HILDEN, S. NEGAHDARIPOUR - *Closed-Form Solution of Absolute Orientation Using Orthonormal Matrices.* Journal of Optical Society of America, A5, pp 1127-1135, 1988

[41] S. HRABAR, G. SUKHATME - *Vision-Based Navigation Through Urban Canyon*. Journal of Field Robotics, Vol. 26(5), pp 431-452, 2009

[42] M-D. HUA, T. HAMEL, P. MORIN, C. SAMSON - *Control of Thrust-Propelled Underactuated Vehicles*. INRIA Technical Report 6453, 2008

[43] M-D. HUA, T. HAMEL, P. MORIN, C. SAMSON - *A Control Approach for Thrust-Propelled Underactuated Vehicles and its Application to VTOL Drones*. IEEE Transactions on Automatic Control, Vol 54(8), pp 1837-1853, 2009

[44] J. ISRAEL, A. PLYER - *Brute SLAM.* Accepted in the 1st IEEE Workshop on Consumer Depth Camers for Computer Vision, 2011

[45] M. KAESS, K. NI, F. DELLAERT - *Flow Separation for Fast and Robust Stereo Odometry.* proceedings of IEEE International Conference on Robotics and Automation, 2009

[46] S. LANGE, N. SUNDERHAUF, P. NEUBERT, and al. - *Autonomous Corridor Flight of a UAV Using a Low-Cost and Light-Weight RGB-D Camera*. Proceedings of International Symposium on Autonomous minirobots for research and edutainment, 2011.

[47] G. LE BESNERAIS, F. CHAMPAGNAT, G. ROCHEFORT - *Robust Optical Flow Estimation Using B-Spline Image Models*. Proceedings of International Symposium on Signal Processing and its Applications, 2003

[48] G. LE BESNERAIS, F. CHAMPAGNAT - *Dense Optical Flow Estimation by Iterative Local Window Registration*. Proceedings of IEEE International Conference on Image Processing, 2005

[49] G. LE BESNERAIS, M. SANFOURCHE, F. CHAMPAGNAT - *Dense Height Map Estimation from Oblique Aerial Image Sequences*. Journal of Computer Vision and Image Understanding, Vol. 109(2), pp. 204-225, 2008

[50] F. LE BRAS, T. HAMEL, R. MAHONY, A. TREIL - *Output Feedback Observation and Control for Visual Servoing of VTOL UAVs*. International Journal of Robust and Nonlinear Control, 2010

[51] M. LHUILLIER, L. QUAN - *Match Propagation for Image-Based Modeling and Rendering.* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24(8), pp 1140-1146 , 2002

[52] H.C. LONGUET-HIGGINS, K. PRAZDNY - *The Interpretation of a Moving Retinal Image*. Proceedings of Royal Society of London, 1981

[53] D. LOWE - *Distinctive Image Features form Scale Invariant Keypoints*. International Journal of Computer Vision, Vol. 60, Issue 2, pp 91-110 , 2004.

[54] B. LUCAS, T. KANADE - *An iterative Image Registration Technique with an Application to Stereo Vision.* Proceedings DARPA Image Understanding Workshop, pp. 121-130, April 1981

[55] A. MALLET, S. LACROIX, L. GALLO - *Position Estimation in Outdoor Environments Using Pixel Tracking and Stereovision.* Proceedings of IEEE International Conference on Robotics and Automation, 2000

[56] S.J. MAYBANK - *The angular velocity associated with the Optical Flowfield arising from Motion Through a Rigid Environment.* Proceedings of The Royal Society of London, 1985

[57] N. METNI, T. HAMEL - *A UAV for Bridge's Inspection : Visual Servoing Control Law With Orientation Limits.* Automation in Construction, 2007

[58] K. MIKOLAJCZYK, C. SCHMID - *Scale and Affine Invariant Interest Point Detectors.* International Journal of Computer Vision, Vol. 60 (1), pp 63-86, 2004

[59] R. MOLERO, S. SCHERER, L. CHAMBERLAIN, S. SINGH - *Navigation and Control for Micro Aerial Vehicles in GPS-Denied Environments*. Technical Report, CMU-RI-TR-10-08, 2010

[60] M. MONTEMERLO, S. THRUN, D. KOLLER, B. WEGBREIT - *FastSLAM 2.0 : An improved Particle Filtering Algorithm for Simultaneous Localization and Mapping that Provably Converges*. International Joint Conference on Artificial Intelligence, Vol. 18, pp 1151-1156, 2003

[61] H.P. MORAVEC - *Robot Spatial Perception by Stereoscopic Vision and 3D Evidence Grids.* MIT technical Report XXX, 1996

[62] E. MOURAGNON, M. LHUILLIER, M. DHOME, F. DEKEYSER, P. SAYD - *Real-Time Localization and 3D Reconstruction.* Proceedings of IEEE conference on Computer Vision and Pattern Recognition, 2006

[63] D. NISTÉR - *An Efficient Solution to the Five-Point Relative Pose Problem.* Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2003

[64] D. NISTÈR, O. NARODITSKY, J. BERGEN - *Visual Odometry for Ground Vehicle Applications.* Journal of Field Robotics, Vol. 23(1), pp. 3-20, 2006

[65] J.M. ODOBEZ, P. BOUTHEMY - *Detection of Multiple Moving Objects Using Multiscale MRF with Camera Motion Compensation*. Proceedings of 1st IEEE International Conference on Image Processing, 1994

[66] L. PAZ, J. TARDOS, J. NEIRA - *Divide and Conquer : EKF SLAM in O(n).* IEEE Transactions on Robotics, Vol 24(5), pp 1107-1120, 2008

[67] B.V. PHAM, S. LACROIX, M. DEVY - *Vision-Based Absolute Navigation for Descent and Landing.* Journal of Field Robotics, 2012

[68] P. PINIES, J. TARDOS - *Large scale SLAM Building Conditionally Independent Local Maps: Application to Monocular Vision.* IEEE Transactions on Robotics, Vol 24(5), pp 1094-1106, 2008

[69] Pix4D - http://pix4d.com/

[70] H. de PLINVAL, P. MORIN, P. MOUYON, T. HAMEL - *Visual Servoing for Underactuated VTOL UAVs : a Linear Homography-Based Approach*. Proceedings of IEEE International Conference on Robotics and Automation, 2011

[71] H. de PLINVAL, P. MORIN, P. MOUYON - *Nonlinear Control of Underactuated Vehicles with Uncertain Position Measurements and Application to Visual Servoing for Underactuated VTOL UAVs*. Submitted to the IEEE American Control Conference, 2012

[72] A. PLYER, G. LE BESNERAIS, F. CHAMPAGNAT - FOLKI GPU : http://www.onera.fr/dtim-en/gpu-for-image/folkigpu.php

[73] S. RAMALINGAM, P. STURM, S. LODHA - *Towards Complete Generic Camera Calibration*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005

[74] E. ROSTEN, T. DRUMMOND - *Machine Learning for High-Speed Corner Detection*. Proceedings of the 9th European Conference on Computer Vision, 2006

[75] J.P. ROUSSEEUW - *Least Median of Squares Regression.* Journal of American Statistical Association, Vol. 79 (388), pp 871-880, 1984

[76] C. ROUSSILLON, A. GONZALEZ, J. SOLA, and al. - *RT-SLAM : A Generic and Real-Time visual SLAM Implementation*. Proceedings of International Conference on Computer Vision Systems, 2011

[77] S. SCHERER, S. SINGH, L. CHAMBERLAIN, M. ELGERSMA - *Flying Low and Fast Among Obstacles: Methodology and Experiments.* Internal Journal of Robotics Research, Vol. 27(5), pp 549-574, 2008

[78] G. SIBLEY, L. MATTHIES, G. SUKHATME - *A Sliding Window Filter for Incremental SLAM*. Unifying Perspectives in Computational and Robot Vision, pp 103-112, 2008

[79] G. SIBLEY, C. MEI, I. REID, P. NEWMAN - *Vast Scale Outdoor Navigation Using Adaptive Relative Bundle Adjustment*. International Journal of Robotics Research, Vol. 29(8), pp 958-980, 2010

[80] J. SHI, C. TOMASI - *Good Features to Track*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1994

[81] M. SRINIVASAN, S. ZHANG, M. LEHRER, T. COLLETT - *Honeybee Navigation en Route to the Goal : Visual Flight Control and Odometry.* Journal of Experimental Biology, Vol 199, pp 237–244, 1996

[82] P. STURM, Z. CHENG, P. C. YU CHEN, A. NEOW POO - *Focal Length Calibration from two Views: Method and Analysis of Singular Cases*. Computer Vision and Image Understanding, Vol. 99(1), pp. 58-95, 2005

[83] E. TOLA, V. LEPETIT, P. FUA - *A Fast Local Descriptor for Dense Matching*. Proceedings of IEEE conference on Computer Vision and Pattern Recognition, 2008

[84] B. TRIGGS, P. MCLAUCHLAN, R. HARTLEY, A. FITZGIBBON - *Bundle Adjustment, a Modern Synthesis*. International Workshop on Vision Algorithms during ICCV, pp 298-372,1999

[85] P. TROUVÉ, F. CHAMPAGNAT, G. LE BESNERAIS, J. IDIER - *Single Image Local Blur Identification*. Proceedings of IEEE International Conference on Image Processing, 2011

[86] R.F. VASSALLO, J. SANTOS-VICTOR, H.J. SCHNEEBELI - *A General Approach for Egomotion Estimation with Omnidirectional Images*. Proceedings of Omnivis'02 Workshop on Omni-directional Vision, 2002

[87] P. VIOLA, W.M. WELLS III - *Alignment by maximization of Mutual Information*. International Journal of Computer Vision, Vol. 24(2), pp 137–154, 1997

[88] Y. WATANABE, P. FABIANI, G. LE BESNERAIS - *Simultaneous Visual Target Tracking and Navigation in a GPS-Denied Environment*. Proceedings of International Conference on Advanced Robotics, 2009

[89] Y. WATANABE and al. - *AZUR : navigation Autonome en Zone URbaine*, Onera research project. Website: http://sites.onera.fr/azur

[90] S.WEISS, D. SCARAMUZZA, R. SIEGWART - *Monocular-SLAM-Based Navigation for Autonomous Micro Helicopters in GPS-Denied Environments.* Journal of Field Robotics, Vol 28(6), pp 854-874, 2011

[91] S. WEISS, M. ACHTELIK, L. KNEIP, D. SCARAMUZZA, R. SIEGWART - *Intuitive 3D Maps for MAV Terrain Exploration and Obstacle Avoidance*. Journal of Intelligent Robotics systems, Vol. 61, pp 473-493,2011

[92] M. WERLBERGER, W. TROBIN, T. POCK, A. WEDEL, D. CREMERS, H. BISCHOF - *Anisotropic Huber-L1 Optical Flow*. Proceedings of British Machine Vision Conference, 2009

[93] K.M. WURM, A. HOMUNG, M. BENNEWITZ, C. STACHNIS, W. BURGARD - *OctoMap: A Probabilistic, Flexible and Compact 3D Map Representation for Robotic Systems*. Proceedings of the IEEE International Conference on Robotics and Automation, 2010

Martial Sanfourche obtained his M.Sc. in Computer Science while at the University of Cergy-Pontoise in 2001 and then a Ph.D. degree in image and signal processing from the University of Cergy-Pontoise in 2005. After a postdoctoral position at CNRS-LAAS he joined Onera-DTIM in 2007 where is now a research engineer in computer vision. His current research interest includes online and offline visual localization and mapping.

Philippe Cornic has been working at Onera as a research engineer for twenty years. He is interested in computer vision and more specifically in matching and image registration.

Jeff Delaune graduated from the Ecole Centrale de Nantes in 2009. He has an M.Sc. in Astronautics and Space Engineering obtained at Cranfield University in the United Kingdom. Since 2010 he has been working as a PhD student at Onera on vision-based navigation for pinpoint planetary landing. He works on tight inertial/vision fusion schemes.

Jonathan Israel graduated from the Ecole Nationale Supérieure des Télécommunications in 2004 and received his Masters degree in mathematics from the Ecole Normale Supérieure, Cachan, France, in 2005. Since 2006 he has been working as a research scientist at Onera. His main interests are related to data registration, segmentation and classification for navigation or interpretation purposes.

Guy Le Besnerais graduated from the Ecole Nationale Superieure de Techniques Avancees in 1989 and received his Ph.D. degree in physics from the Universite de Paris-Sud, Orsay, France, in 1993. He joined Onera in 1994, where he is now a senior scientist in the Information Processing and Modelization Department. His work concerns inversion problems in imagery and computer vision, with an emphasis on embedded perception for autonomous aerial robots.

Henry de Plinval obtained his M.Sc. in Aeronautics and Astronautics at MIT and his Engineer diploma at the Ecole Polytechnique in 2006. He is working as a researcher at Onera, the French Aerospace lab. His current research interests include navigation, guidance and control of systems, especially Unmanned Aerial Vehicles.

Aurelien Plyer (DTIM) is a Onera-LAGA PhD student since 2008. He received his M.Sc. in Computer Science from the Université Pierre et Marie Curie (Paris 6) in 2008. His research deals with video analysis and interpretation for aerial videos within a urban context and he is using GPU programming in order to implement real-time processing.

Aurélie Treil has been a Onera-DCSD PhD student since 2009. She received her Master of Science in Automatics and signal treatment from Supélec/Université Paris 6 and obtained an Engineering degree at ESTACA in 2009. Her research deals with UAV's visual servoing.

Yoko Watanabe has been working at Onera-DCSD as a research engineer since 2008. She received her Master's degree in Aeronautics and Astronautics from Kyoto University (Japan) in 2003, and her Ph.D. in Aerospace Engineering from the Georgia Institute of Technology (USA) in 2008. Her research interests include navigation, guidance and control of autonomous robots. She has, in particular, a lot of experience of vision-based navigation and the guidance of unmanned aerial vehicles.