J Intell Robot Syst
DOI 10.1007/s10846-008-9257-y

# Vision-Based Odometry and SLAM for Medium and High Altitude Flying UAVs

**F. Caballero · L. Merino · J. Ferruz · A. Ollero**

Received: 15 March 2008 / Accepted: 30 June 2008


**Abstract** This paper proposes vision-based techniques for localizing an unmanned aerial vehicle (UAV) by means of an on-board camera. Only natural landmarks provided by a feature tracking algorithm will be considered, without the help of visual beacons or landmarks with known positions. First, it is described a monocular visual odometer which could be used as a backup system when the accuracy of GPS is reduced to critical levels. Homography-based techniques are used to compute the UAV relative translation and rotation by means of the images gathered by an onboard camera. The analysis of the problem takes into account the stochastic nature of the estimation and practical implementation issues. The visual odometer is then integrated into a simultaneous localization and mapping (SLAM) scheme in order to reduce the impact of cumulative errors in odometry-based position estimation

This work is partially supported by the AWARE project (IST-2006-33579) funded by the European Commission, and the AEROSENS project (DPI-2005-02293) funded by the Spanish Government.

F. Caballero (✉) · J. Ferruz · A. Ollero
University of Seville, Seville, Spain
e-mail: caba@cartuja.us.es

J. Ferruz
e-mail: ferruz@cartuja.us.es

A. Ollero
e-mail: aollero@cartuja.us.es

L. Merino
Pablo de Olavide University, Seville, Spain
e-mail: lmercab@upo.es

approaches. Novel prediction and landmark initialization for SLAM in UAVs are presented. The paper is supported by an extensive experimental work where the proposed algorithms have been tested and validated using real UAVs.

**Keywords** Visual odometry · Homography · Unmanned aerial vehicles · Simultaneous localization and mapping · Computer vision

## 1 Introduction

Outdoor robotics applications in natural environments sometimes require different accessibility capabilities than the capabilities provided by existing ground robotic vehicles. In fact, in spite of the progress in the development of unmanned ground vehicles along the last 20 years, navigating in unstructured natural environments still poses significant challenges. The existing ground vehicles have inherent limitations to reach the desired locations in many applications. The characteristics of the terrain and the presence of obstacles, together with the requirement of fast response, may represent a major drawback to the use of any ground locomotion system. Thus, in many cases, the use unmanned aerial vehicles (UAVs) is the only effective way to reach the target to get information or to deploy instrumentation.

In the last ten years UAVs have improved their autonomy both in energy and information processing. Significant achievements have been obtained in autonomous positioning and tracking. These improvements are based on modern satellite-based position technologies, inertial navigation systems, communication and control technologies, and image processing. Furthermore, new sensing and processing capabilities have been implemented on-board the UAVs. Thus, today we can consider some UAVs as intelligent robotic systems integrating perception, learning, real-time control, situation assessment, reasoning, decision-making and planning capabilities for evolving and operating in complex environments.

In most cases, UAVs use the global position system (GPS) to determine their position. As pointed out in the Volpe Report [42], the accuracy of this estimation directly depends on the number of satellites used to compute the position and the quality of the signals received by the device; radio effects like multi-path propagation could cause the degradation in the estimation. In addition, radio frequency interferences with coexisting devices or jamming could make the position estimation unfeasible.

These problems are well known in robotics. Thus, odometry is commonly used in terrestrial robots as a backup positioning system or in sensor data fusion approaches. This local estimation allows temporally managing GPS faults or degradations. However, the lack of odometry systems in most aerial vehicles can lead to catastrophic consequences under GPS errors; incoherent control actions could be commanded to the UAV, leading to crash and the loss of valuable hardware. Moreover, if full autonomy in GPS-less environments is considered, then the problem of simultaneous localization and mapping (SLAM) should be addressed.

If small UAVs are considered, their low payload represents a hard restriction on the variety of devices to be used for odometry. Sensors like 3D or 2D laser scanners

are too heavy and have an important dependence to the UAV distance to the ground. Although there exist small devices for depth sensing, their range is usually shorter than 15 m. Stereo vision systems have been successfully applied to low/medium size UAVs due to its low weight and versatility [4, 9, 18], but the rigid distance between the two cameras limits the useful altitude range.

Monocular vision seems to offer a good solution in terms of weight, accuracy and scalability. This paper proposes a monocular visual odometer and vision-based localization methods to act as backup systems when the accuracy of GPS is reduced to critical levels. The objective is the development of computer vision techniques for the computation of the relative translation and rotation, and for the localization of the vehicle based on the images gathered by a camera on-board the UAV. The analysis of the problem takes into account the stochastic nature of the estimation and practical implementation issues.

The paper is structured as follows. First, related work in vision based localization for UAVs is detailed. Then, a visual odometer based on frame-to-frame homographies is described, together with a robust method for homography computation. Later, the homography-based odometry is included in a SLAM scheme in order to overcome the error accumulation present in odometric approaches. The proposed SLAM approach uses the information provided by the odometer as main prediction hypothesis and for landmark initialization. Finally, conclusions and lessons learned are described.

### 1.1 Related Work

One of the first researches on vision applied to UAV position estimation starts in the nineties at the Carnegie-Mellon University (CMU). In [1], it is described a vision-based odometer that allowed to lock the UAV to ground objects and sense relative helicopter position and velocity in real time by means of stereo vision. The same visual tracking techniques, combined with inertial sensors, were applied to autonomous take off, following a prescribed trajectory and landing. The CMU autonomous helicopter also demonstrated autonomous tracking capabilities of moving objects by using only on-board specialized hardware.

The topic of vision-based autonomous landing of airborne systems has been actively researched [30]. In the early nineties, Dickmanns and Schell [13] presented some results of the possible use of vision for landing an airplane. Systems based on artificial beacons and structured light are presented [44, 45]. The BEAR project at Berkeley is a good example of vision systems for autonomous landing of UAVs. In this project, vision-based pose estimation relative to a planar landing target and vision-based landing of an aerial vehicle on a moving deck have been researched [36, 40]. A technique based on multiple view geometry is used to compute the real motion of one UAV with respect to a planar landing target. An artificial target allows to establish quick matches and to solve the scale problem.

Computer vision has also been proposed for safe landing. Thus, in [15], a strategy and an algorithm relying on image processing to search the ground for a safe landing spot is presented. Vision-based techniques for landing on a artificial helipad of known shape are also presented in [34, 35], where the case of landing on a slow moving

helipad is considered. In [37], the landing strategies of bees are used to devise a vision system based on optical flow for UAVs.

Corke et. al [9] have analyzed the use of stereo vision for height estimation in small size helicopters. In Georgia Tech, vision-based aided navigation for UAVs has been considered. Thus, in [43] the authors present an Extended Kalman Filter approach that combines GPS measurements with image features obtained from a known artificial target for helicopter position estimation.

In a previous work [5], the authors present a visual odometer for aerial vehicles using monocular image sequences, but no error estimation is provided by the algorithm, and the approach is limited to planar scenes. In [6], it is shown how a mosaic can be used in aerial vehicles to partially correct the drift associated to odometric approaches. This technique is extended in [7] with a minimization process that allows to improve the spatial consistency of the online built mosaic. Recently, in [8] the authors propose a visual odometer to compensate GPS failures. Image matching with geo-referenced aerial imagery is proposed to compensate the drift associated to odometry.

Although vision-based SLAM has been widely used in ground robots and has demonstrated its feasibility for consistent perception of the environment and position of the robot, only a few applications have been implemented on UAVs. The researches carried out in the LAAS laboratory in France and the Centre for Autonomous Systems in Australia can be highlighted. The first of them has developed an stereo vision system designed for the KARMA blimp [18, 21], where interest point matching and Kalman filtering techniques are used for simultaneous localization and mapping with very good results. However, this approach is not suitable for helicopters, as the baseline of the stereo rig that can be carried is small, and therefore it limits the height at which the UAV can fly. UAV simultaneous localisation and map building with vision using a delta fixed-wing platform is also presented in [19]. Artificial landmarks of known size are used in order to simplify the landmark identification problem. The known size of the landmarks allows to use the cameras as a passive range/bearing/elevation sensor. Preliminary work on the use of vision-based bearing-only SLAM in UAVs is presented in [23]. In [22], vision and IMU are combined for UAV SLAM employing an Unscented Kalman Filter. The feature initialization assumes a flat terrain model, similarly to the present approach. Results in simulation are shown in the paper. In [25], an architecture for multi-vehicle SLAM is studied for its use with UAVs. The paper deals with the issues of data association and communication, and some simulation results are presented.

Visual servoing approaches has been also proposed for direct control of UAVs. The use of an omnidirectional camera for helicopter control has been presented in [17]. The camera is used to maintain the helicopter in the centroid of a set of artificial targets. The processed images are directly used to command the helicopter. The paper shows the feasibility of the procedure, but no actual control is tested. Omnidirectional vision is also used in [12] to estimate the attitude of an UAV. The method detects the horizon line by means of image processing and computes the attitude from its apparent motion. In the work of [27], vision is used to track features of buildings. Image features and GPS measurements are combined together to keep the UAV aligned with the selected features. Control design and stability analysis of image-based controllers for aerial robots are presented in [26]. In [32] recent work on vision-based control of a fixed wing aircraft is presented.

## 2 Homography-Based Visual Odometry for UAVs

Image homographies will be a basic tool for estimating the motion that an UAV undergoes by using monocular image sequences. A homography can be defined as an invertible application of the projective space $\mathbf{P^2}$ into $\mathbf{P^2}$ that applies lines into lines. Some basic properties of the homographies are the following:

– Any homography can be represented as a linear and invertible transformation in homogeneous coordinates:

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ k \end{bmatrix} = \underbrace{\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{1}$$

  Inversely, any transformation of this nature can be considered as a homography.
– Given the homogeneous nature of the homography **H**, it can be multiplied by an arbitrary constant $k \neq 0$ and represent the same transformation. This means that the matrix **H** is constrained by eight independent parameters and a scale factor.

Given two views of a scene, the homography model represents the exact transformation of the pixels on the image plane if both views are related by a pure rotation, or if the viewed points lie on a plane. When a UAV flies at relatively high altitude, it is a usual assumption to model the scene as pseudo-planar. The paper will propose a method to extend the applicability of the homography model to non-planar scenes (computing the homography related to a dominant plane on the scene) in order to be able to perform motion estimation at medium or even low UAV altitude.

### 2.1 Robust Homography Estimation

The algorithm for homography computation is based on a point features matching algorithm, and has been tested and validated with thousands of images captured by different UAVs flying at different altitudes, from 15 to 150 m. This algorithm (including the feature matching approach) was briefly described in [29]. It basically consists of a point-feature tracker that obtains matches between images, and a combination of least median of squares and M-Estimator for outlier rejection and accurate homography estimation from these matches.

However, there are two factors that may reduce the applicability of the technique, mainly when the UAV flies at altitudes of the same order of other elements on the ground (buildings, trees, etc):

– Depending on the frame-rate and the vehicle motion, the overlap between images in the sequence is sometimes reduced. This generates a non-uniform distribution of the features along the images.
– In 3D scenes, the parallax effect will increase, and the planarity assumption will not hold. The result is a dramatic growth of the outliers and even the divergence of the M-Estimator.

They produce different problems when computing the homography. If the matches are not uniformly distributed over the images, an ill-posed system of equations for homography computation will be generated, and there may exist multiple solutions. On the other hand, if the parallax effect is significant, there may exist multiple planes (whose transformation should be described by multiple homographies); the algorithm should try to filter out all features but those lying on the dominant plane of the scene (the ground plane).

In the proposed solution, the first problem is addressed through a *hierarchy* of homographic models (see Fig. 1), in which the complexity of the model to be fitted is decreased whenever the system of equations is ill-constrained, while the second is tackled through outlier rejection techniques.

Therefore, depending on the quality of the available data, the constraints used to compute the homography are different; thus, the accuracy changes as well. An estimation of this accuracy will be given by the covariance matrix of the computed parameters.

A complete homography has 8 *df* (as it is defined up to a scale factor). The degrees of freedom can be reduced by fixing some of the parameters of the $3 \times 3$ matrix. The models used are the defined by Hartley in [16]: Euclidean, Affine and Complete Homographic models, which have 4, 6 and 8 *df* respectively (see Fig. 1). The percentage of successful matches obtained by the point tracker is used to have an estimation about the level of the hierarchy where the homography computation should start. These percentage thresholds were obtained empirically by processing hundreds of aerial images. Each level involves the following different steps:

- Complete homography. Least median of squares (LMedS) is used for outlier rejection and a M-Estimator to compute the final result. This model is used if more than the 65% of the matches are successfully tracked.
- Affine homography. If the percentage of success in the tracking step is between 40% and 65%, then the LMedS is not used, given the reduction in the number of matches. A relaxed M-Estimator (soft penalization) is carried out to compute the model.

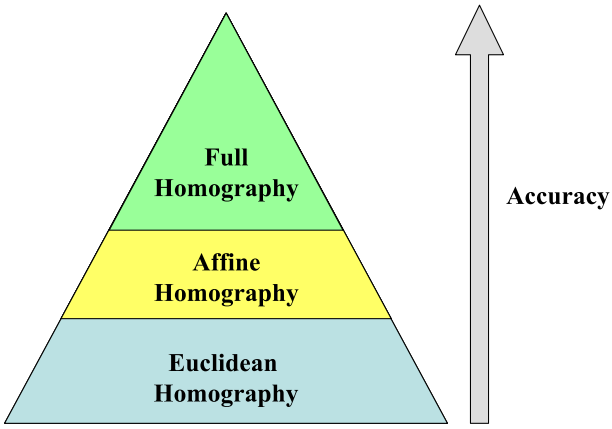


**Fig. 1** Levels in the proposed hierarchical homography computation. Accuracy increases with the complexity of the model

– Euclidean homography. If the percentage is below 40%, the set of data is too noisy and small to apply non-linear minimizations. The model is computed using least-squares.

In addition, it is necessary a rule to know when the current level is ill-posed and the algorithm has to decrease the model complexity. The M-Estimator used in the complete and affine computations is used for this purpose. It is considered that the M-Estimator diverge if it reaches the maximum number of iterations and, hence, the level in the hierarchy has to be changed to the next one.

## 2.2 Geometry of Two Views of the Same Plane

The odometer will extract the camera motion from the image motion modeled by the estimated homography between two consecutive views. If we consider the position and orientation of two cameras in the world coordinate frame, as shown in Fig. 2, it can be seen that the two projections $\mathbf{m_1} \in \mathbb{R}^2$ and $\mathbf{m_2} \in \mathbb{R}^2$ of a fixed point $\mathbf{P} \in \mathbb{R}^3$ belonging to a plane $\Pi$ are related by:

$$\tilde{\mathbf{m}}_2 = \underbrace{\mathbf{A}_2 \mathbf{R}_{12} \left( \mathbf{I} - \frac{\mathbf{t}_2 \mathbf{n}_1^T}{d_1} \right) \mathbf{A}_1^{-1}}_{\mathbf{H}_{12}} \tilde{\mathbf{m}}_1 \tag{2}$$

where $\mathbf{R}_{12}$ is the rotation matrix that transforms a vector expressed in the coordinate frame of camera one into the coordinate frame of camera two, $\mathbf{t}_2$ is the translation of camera two with respect to camera one expressed in the coordinate frame of camera one, the Euclidean distance from the camera one to the plane $\Pi$ is $d_1$ and the normal of the plane $\Pi$ (in the first camera coordinate frame) is given by the unitary 3-D vector $\mathbf{n}_1$ (see Fig. 2).

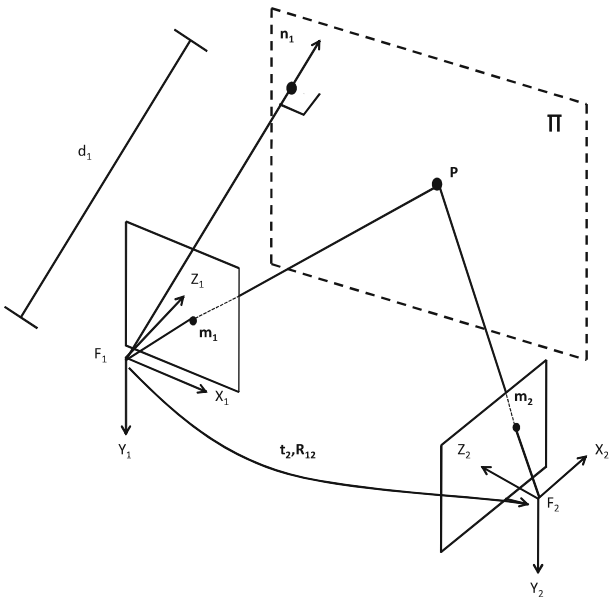


**Fig. 2** Geometry of two views of the same plane

For this particular case, the transformation between the features $\mathbf{m}_1$ and $\mathbf{m}_2$ is a plane-to-plane homography, so $\tilde{\mathbf{m}}_2 = \mathbf{H}_{12}\tilde{\mathbf{m}}_1$. This homography is completely defined by the calibration matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, the relative position of the cameras and the structure of the scene (the normal and distance of the plane). The problem can be reformulated as a single camera whose position and orientation change through time. In this case the calibration matrix is the same for both views, so $\mathbf{A}_1 = \mathbf{A}_2$.

Then, for the calibrated case, the relative position (rotation and translation) between the cameras and the plane normal can be obtained if the homography that relates two views of the same plane is known, for instance by obtaining a set of matches between the images, as described in the previous section. Moreover, it will be shown how to obtain an estimation of the covariance matrix for all these parameters.

### 2.3 Motion Estimation from Homographies

A solution based on the singular value decomposition (SVD) of the homography will be used. Consider a single camera that moves through time, the homography $\mathbf{H}_{12}$ that relates the first and the second view of the same planar scene and the camera calibration matrix $\mathbf{A}_1$. According to Eq. 2, the *calibrated homography* is defined as:

$$\mathbf{H}_{12}^{c} = \mathbf{A}_1^{-1}\mathbf{H}_{12}\mathbf{A}_1 = \mathbf{R}_{12}\left(\mathbf{I} - \frac{\mathbf{t}_2\mathbf{n}_1^T}{d_1}\right) \tag{3}$$

The elements can be extracted from the singular value decomposition (SVD) of the homography $\mathbf{H}_{12}^{c} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, where $\mathbf{D} = diag(\lambda_1, \lambda_2, \lambda_3)$ stores the singular values. Once $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$ have been conveniently ordered such us $\lambda_1 > \lambda_2 > \lambda_3$, the singular values can be used to distinguish three types of movements carried out by the camera [39]:

- The three singular values of $\mathbf{H}_{12}^{c}$ are equal, so $\lambda_1 = \lambda_2 = \lambda_3$. It occurs when the motion consist of rotation around an axis through the origin only, i.e., $\mathbf{t}_2 = \mathbf{0}$. The rotation matrix is unique, but there is not sufficient information to estimate the plane normal $\mathbf{n}_1$.
- The multiplicity of the singular values of $\mathbf{H}_{12}^{c}$ is two, for example $\lambda_1 = \lambda_2 \neq \lambda_3$. Then, the solution for motion and geometrical parameters is unique up to a common scale factor for the translation parameters. In this case, the camera translation is parallel to the normal plane.
- The three singular values of $\mathbf{H}_{12}^{c}$ are different, i.e., $\lambda_1 \neq \lambda_2 \neq \lambda_3$. In this case two possible solutions for rotation, translation and plane normal exist and can be computed.

The presence on noise in both feature tracking and homography estimation always leads to different singular values for $\mathbf{H}_{12}^{c}$ and the third of the previous cases becomes

the dominant in real conditions. Rotation, translation and normal to the plane is then given by the following expressions [39]:

$$\begin{aligned} \mathbf{R}_2 &= \mathbf{U}\begin{bmatrix} \alpha & 0 & \beta \\ 0 & 1 & 0 \\ -s\beta & 0 & s\alpha \end{bmatrix}\mathbf{V}^T \\ \mathbf{t}_2 &= \frac{1}{w}\left(-\beta\mathbf{u}_1 + \left(\frac{\lambda_3}{\lambda_2} - s\alpha\right)\mathbf{u}_3\right) \\ \mathbf{n}_1 &= w(\delta\mathbf{v}_1 + \mathbf{v}_3) \end{aligned} \tag{4}$$

where:

$$\begin{aligned} \delta &= \pm\sqrt{\frac{\lambda_1^2 - \lambda_2^2}{\lambda_2^2 - \lambda_3^2}} \\ \alpha &= \frac{\lambda_1 + s\lambda_3\delta^2}{\lambda_2(1+\delta^2)} \\ \beta &= \pm\sqrt{1-\alpha^2} \\ s &= det(\mathbf{U})det(\mathbf{V}) \end{aligned}$$

and $\omega$ is a scale factor. We set that scale factor so that $\|\mathbf{n}_1\| = 1$. Each solution must accomplish that $sgn(\beta) = -sgn(\delta)$. For this case, Triggs algorithm [38] allows a systematic and robust estimation. This method has been implemented and tested in the experiments presented in this paper with very good results.

From Eq. 3 it can be seen that (as $\|\mathbf{n}_1\| = 1$) only the product $\frac{\|\mathbf{t}_2\|}{d_1}$ can be recovered. The scale can be solved, then, if the distance $d_1$ of camera 1 to the reference plane is known. If the reference plane is the ground plane, as it would be the case in the experiments, a barometric sensor or height sensor can be used to estimate this initial distance. Also, a range sensor can be used. In this paper, we will consider that this height is estimated for the first frame by one of these methods.

### 2.4 Correct Solution Disambiguation

Apart from the scale factor, two possible solutions $\{\mathbf{R}_2^1, \mathbf{t}_2^1, \mathbf{n}_1^1\}$ and $\{\mathbf{R}_2^2, \mathbf{t}_2^2, \mathbf{n}_1^2\}$ will be obtained. Given a third view and its homography with respect to the first frame $\mathbf{H}_{13}$, it is possible to recover an unique solution, as the estimated normal of the reference plane in the first camera coordinate frame, $\mathbf{n}_1$, should be the same.

A method to detect the correct solution is proposed. If a sequence of images is used, the set of possible normals is represented by:

$$S_n = \{\mathbf{n}_{12}^1, \mathbf{n}_{12}^2, \mathbf{n}_{13}^1, \mathbf{n}_{13}^2, \mathbf{n}_{14}^1, \mathbf{n}_{14}^2, ...\} \tag{5}$$

where the superindex denotes the two possible normal solutions and the subindex $1\,j$ denotes the normal $\mathbf{n}_1$ estimated using image $j$ in the sequence.

If $\mathbf{n}_{12}^1$ and $\mathbf{n}_{12}^2$ were correct, there would be two set of solutions, $S_{n^1}$ and $S_{n^2}$. The uniqueness of the normal leads to the following constraints:

$$\left\| \mathbf{n}_{12}^1 - \mathbf{n}_{1j}^i \right\| \leq \epsilon_1 \ \forall \mathbf{n}_{1j}^i \in S_{n^1} \tag{6}$$

$$\left\| \mathbf{n}_{12}^2 - \mathbf{n}_{1j}^i \right\| \leq \epsilon_2 \ \forall \mathbf{n}_{j}^i \in S_{n^2} \tag{7}$$

where $\epsilon_1$ and $\epsilon_2$ are the minimal values that guarantee an unique solution for Eqs. 6 and 7 respectively. The pairs $\{S_{n^1}, \epsilon_1\}$ and $\{S_{n^2}, \epsilon_2\}$ are computed separately by means of the following iterative algorithm:

1. The distance among $\mathbf{n}_{12}^i$ and the rest of normals of $S_n$ is computed.
2. $\epsilon_i$ is set to an initial value.
3. For the current value $\epsilon_i$, check if there exist an unique solution.
4. If no solution is found, increase the value of $\epsilon_i$ and try again with the step 3. If multiple solutions were found decrease $\epsilon_i$ and try again with step 3. If an unique solution was found, then finish.

The algorithm is applied to $i = 1$ and $i = 2$ and the correct solution is then chosen between both options as the one that achieves the minimum $\epsilon$.

### 2.5 An Estimation of the Uncertainties

An important issue with odometric measurements is to obtain a correct estimation of the associated drift. The idea is to estimate the uncertainties on the estimated rotation, translation and plane normal from the covariance matrix associated to the homography, which can be computed from the estimated errors on the point matches [6].

The proposed method computes the Jacobian of the complete process to obtain a first order approximation of rotation, translation and plane normal error covariance matrix. Once the calibrated homography has been decomposed into its singular values, the computation of the camera motion is straightforward, so this section will focus in the computation of the Jacobian associated to the singular value decomposition process.

Thus, given the SVD decomposition of the calibrated homography $\mathbf{H}_{12}^c$:

$$\mathbf{H}_{12}^c = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} = \mathbf{U}\mathbf{D}\mathbf{V}^T = \sum_{i=1}^{3} \left( \lambda_i \mathbf{u}_i \mathbf{v}_i^T \right) \tag{8}$$

The goal is to compute $\frac{\partial \mathbf{U}}{\partial h_{ij}}$, $\frac{\partial \mathbf{V}}{\partial h_{ij}}$ and $\frac{\partial \mathbf{D}}{\partial h_{ij}}$ for all $h_{ij}$ in $\mathbf{H}_{12}^c$. This Jacobian can be easily computed through the robust method proposed by Papadopoulo and Lourakis in [31].

Taking the derivative of Eq. 8 with respect to $h_{ij}$ yields the following expression:

$$\frac{\partial \mathbf{H}_{12}^c}{\partial h_{ij}} = \frac{\partial \mathbf{U}}{\partial h_{ij}} \mathbf{D}\mathbf{V}^T + \mathbf{U}\frac{\partial \mathbf{D}}{\partial h_{ij}} \mathbf{V}^T + \mathbf{U}\mathbf{D}\frac{\partial \mathbf{V}^T}{\partial h_{ij}} \tag{9}$$

Clearly, $\forall(k,l) \neq (i,j)$, $\frac{\partial h_{kl}}{\partial h_{ij}} = 0$ while $\frac{\partial h_{ij}}{\partial h_{ij}} = 1$. Since $\mathbf{U}$ is an orthogonal matrix:

$$\mathbf{U}^T\mathbf{U} = \mathbf{I} \Rightarrow \frac{\partial \mathbf{U}^T}{\partial h_{ij}}\mathbf{U} + \mathbf{U}^T\frac{\partial \mathbf{U}}{\partial h_{ij}} = {\Omega_U^{ij}}^T + \Omega_U^{ij} = 0 \tag{10}$$

where $\Omega_U^{ij}$ is defined by

$$\Omega_U^{ij} = \mathbf{U}^T\frac{\partial \mathbf{U}}{\partial h_{ij}} \tag{11}$$

It is clear that $\Omega_U^{ij}$ is an antisymmetric matrix. Similarly, an antisymmetric matrix $\Omega_V^{ij}$ can be defined for $\mathbf{V}$ as:

$$\Omega_V^{ij} = \frac{\partial \mathbf{V}^T}{\partial h_{ij}}\mathbf{V} \tag{12}$$

By multiplying Eq. 9 by $\mathbf{U}^T$ and $\mathbf{V}$ from left and right respectively, and using Eqs. 11 and 12, the following relation is obtained:

$$\mathbf{U}^T\frac{\partial \mathbf{H_{12}^u}}{\partial h_{ij}}\mathbf{V} = \Omega_U^{ij}\mathbf{D} + \frac{\partial \mathbf{D}}{\partial h_{ij}} + \mathbf{D}\Omega_V^{ij} \tag{13}$$

Since $\Omega_U^{ij}$ and $\Omega_V^{ij}$ are antisymmetric matrices, all their diagonal elements are equal to zero. Recalling that $\mathbf{D}$ is a diagonal matrix, it is easy to see that the diagonal elements of $\Omega_U^{ij}\mathbf{D}$ and $\mathbf{D}\Omega_V^{ij}$ are also zero. Thus:

$$\frac{\partial \lambda_k}{\partial h_{ij}} = u_{ik}v_{jk} \tag{14}$$

Taking into account the antisymmetric property, the elements of the matrices $\Omega_U^{ij}$ and $\Omega_V^{ij}$ can be computed by solving a set of $2 \times 2$ linear systems, which are derived from the off-diagonal elements of the matrices in Eq. 13:

$$\left.\begin{aligned} d_l\Omega_{U\,kl}^{ij} + d_k\Omega_{V\,kl}^{ij} &= \ \ u_{ik}v_{jl} \\ d_k\Omega_{U\,kl}^{ij} + d_l\Omega_{V\,kl}^{ij} &= -u_{il}v_{jk} \end{aligned}\right\} \tag{15}$$

where the index ranges are $k = 1 \ldots 3$ and $l = i+1 \ldots 2$. Note that, since the $d_k$ are positive numbers, this system has a unique solution provided that $d_k \neq d_l$. Assuming for now that $\forall(k,l)$, $d_k \neq d_l$, the 3 parameters defining the non-zero elements of $\Omega_U^{ij}$ and $\Omega_V^{ij}$ can be easily recovered by solving the 3 corresponding $2 \times 2$ linear systems.

Once $\Omega_U^{ij}$ and $\Omega_V^{ij}$ have been computed, the partial derivatives are obtained as follows:

$$\frac{\partial \mathbf{U}}{\partial h_{ij}} = \mathbf{U}\Omega_U^{ij} \tag{16}$$

$$\frac{\partial \mathbf{V}}{\partial h_{ij}} = -\mathbf{V}\Omega_V^{ij} \tag{17}$$

Taking into account the Eqs. 14, 16 and 17 and the covariance matrix corresponding to the homography it is possible to compute the covariance matrix associated to $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$. Further details and demonstrations can be found in [31]. Finally, the

Jacobians of the equations used to extract the rotation, translation and normal, given by Eq. 4, are easily computed and combined with these covariances to estimate the final motion covariances.

### 2.6 Experimental Results

This section shows some experimental results in which the homography-based visual odometer is applied to monocular image sequences gathered by real UAVs.

The first experiment was conducted with the HERO helicopter (see Fig. 3). HERO is an aerial robotic platform designed for research on UAV control, navigation and perception. It has been developed by the "Robotics, Vision and Control Research Group" at the University of Seville during the CROMAT project, funded by the Spanish Government. HERO is equipped with accurate sensors to measure position and orientation, cameras and a PC-104 to allow processing on board. A DSP is used as data acquisition system and low level controller (position and orientation); the PC-104 runs the rest of tasks such as perception, communications or navigation. All the data gathered by the DSP are exported to the PC-104 through a serial line and published for the rest of the processes.

All the sensor data have been logged together with the images in order to avoid inconsistency among different sensor data. The position is estimated with a Novatel DGPS with 2 cm accuracy and updated at 5 Hz, while an inertial measurement unit (IMU) provides the orientation at 50 Hz, with accuracy of 0.5 degrees. In the experiment, the camera was oriented forty-five degrees with respect to the helicopter horizontal.

The visual odometer algorithm (feature tracking, robust homography computation and homography decomposition) has been programmed in C++ code and runs at 10 Hz with $320 \times 240$ images. The experiment image sequence is composed by 650 samples, or approximately 65 s of flight. A sharp movement is made around sample 400.

The DGPS measurements are used to validate the results. Along the flight, good GPS coverage was available at all time. It is important to notice that the

**Fig. 3** HERO helicopter

odometry is computed taking into account the estimated translation and rotation, so it accumulates both errors. The estimated position by using the visual odometer is shown in Fig. 4. The figure presents the DGPS position estimation and the errors associated to the odometer. It can be seen how the errors grow with the image sample index. The errors corresponding to each estimation are added to the previous ones

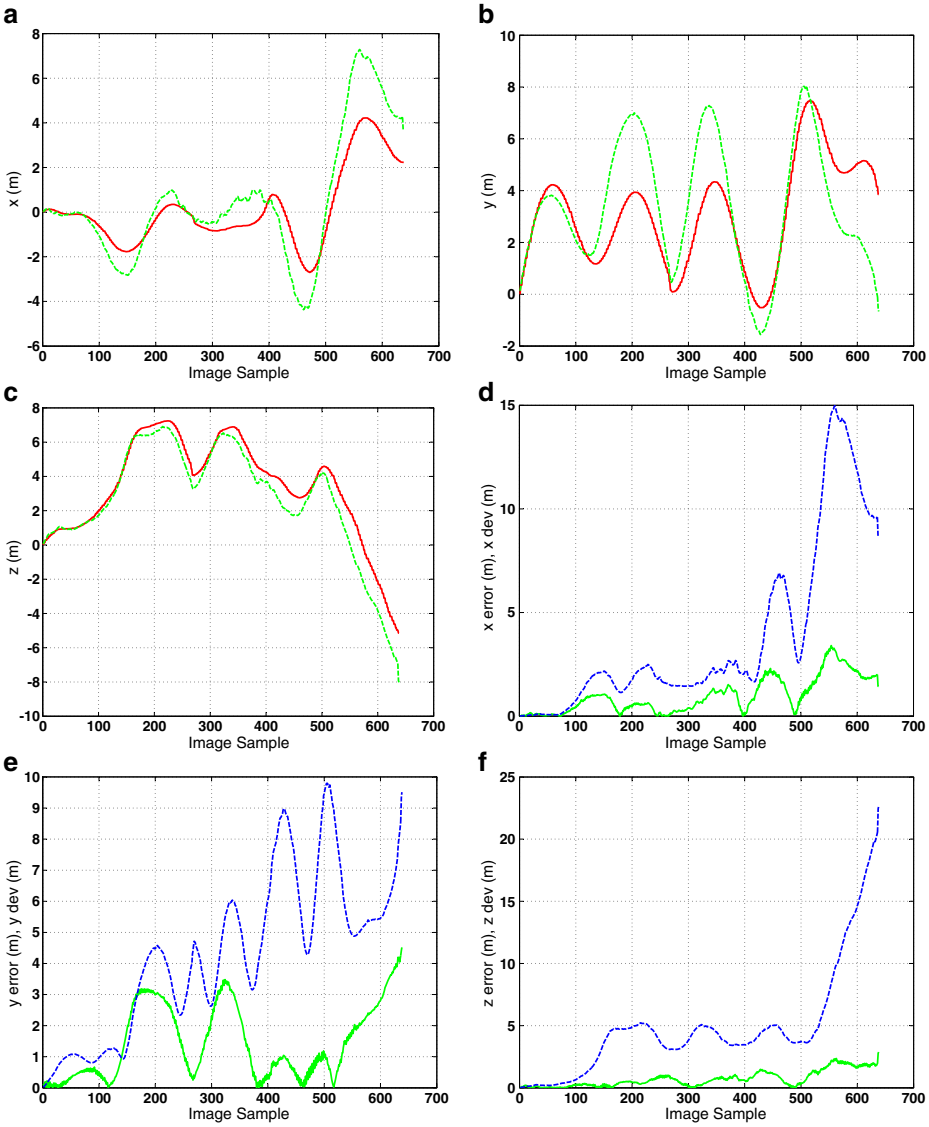


**Fig. 4** *Up* position estimation using vision based technique (*green dashed line*) and DGPS estimation (*red solid line*). *Down* error of the vision based odometry (*green solid line*) and estimated standard deviation (*blue dashed line*)

and make the position estimation diverge through time. Moreover, it can be seen how the estimation of the standard deviation is coherent with the evolution of the error (which is very important for further steps).

Figure 5 shows the evolution of the estimated orientation by using the odometer and the on-board IMU. The orientation has been represented in the classic

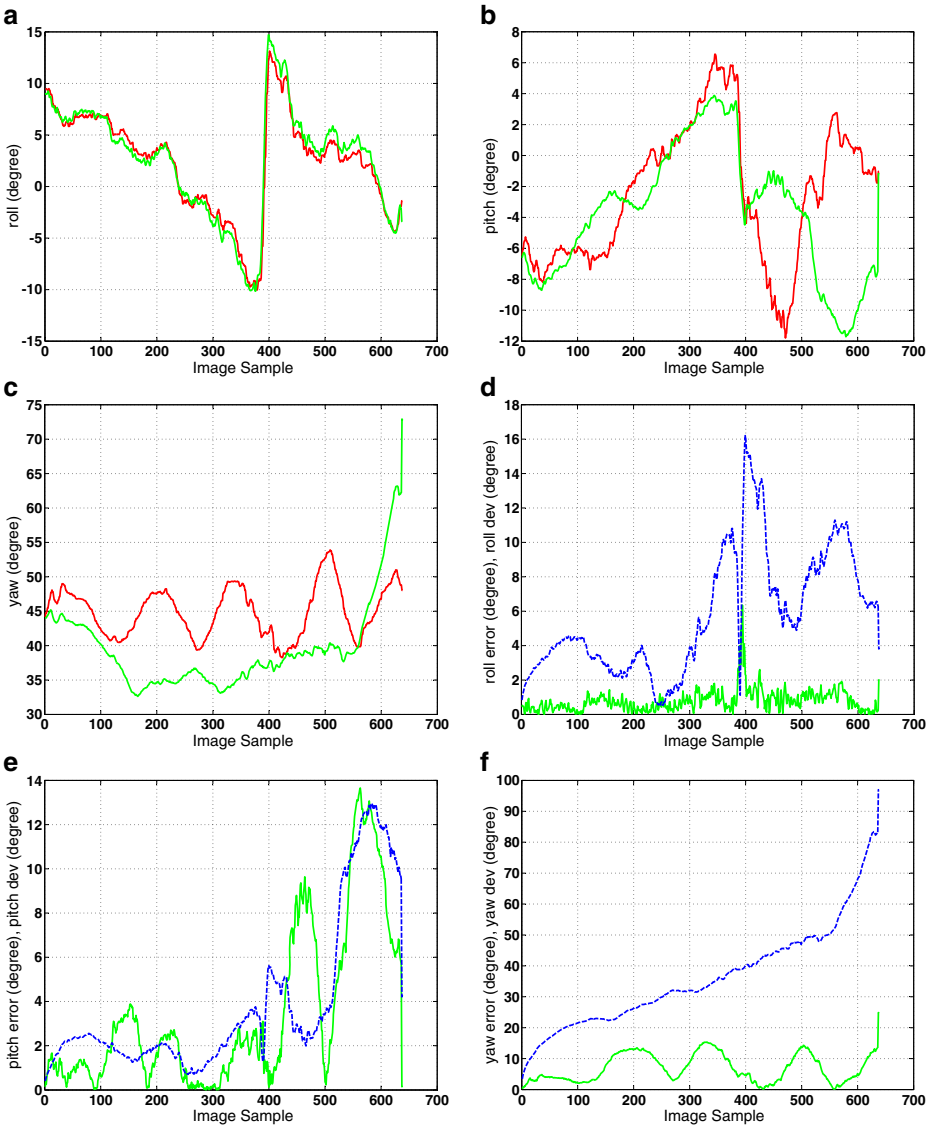


**Fig. 5** *Top* estimated orientation by using vision based technique (*green dashed line*) and IMU estimation (*red solid line*). The orientation is represented in roll/pitch/yaw. *Bottom* errors in the vision based estimation (*green solid line*) and estimated standard deviation (*blue dashed line*)

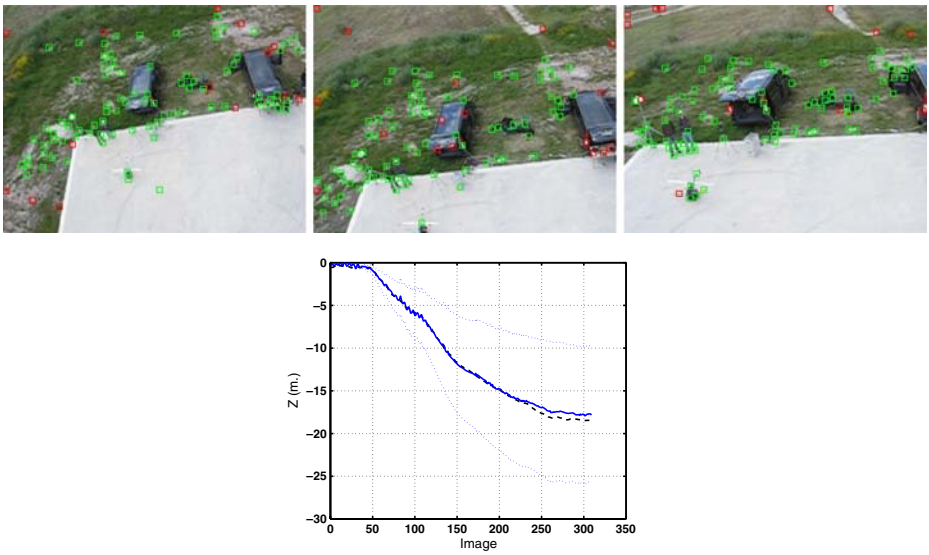


**Fig. 6** Three images of the landing sequence and the estimated height computed by the visual odometer (*solid*) and DGPS (*dashed*). The average frame rate is 7 Hz

roll/pitch/yaw convention (Euler XYZ). It can be seen that the errors in the estimated orientation are small except for the pitch angle. The standard deviation is in general overall consistent.

Results have been also obtained with data gathered during an autonomous landing[1] by the autonomous helicopter Marvin, developed by the Technical University of Berlin [33]. Figure 6 shows three frames of the landing sequence with the obtained matches. It should be pointed out that there are no artificial landmarks for the matching process. Also, in this experiment, the concrete landing platform lacks of structure, which can pose difficulties for the matching procedure. Moreover, along the descent, the pan and tilt unit was moving the camera. Figure 6 shows the estimated translation compared with DGPS, along with the estimated errors. The results are very accurate, although the technique tends to overestimate the uncertainty.

Thus, the experimental results show that the visual odometer can be used to estimate the motion of the UAV; moreover, the estimated errors are consistent. It is important to highlight that all experiments were carried out by using natural landmarks automatically selected by the feature tracking algorithm, without the help of visual beacons.

## 3 Application of Homography-Based Odometry to the SLAM Problem

A SLAM-based technique is proposed to compensate the accumulative error intrinsic to odometry and to solve the localization problem. SLAM employing monocular

[1]The autonomous landing was done based on DGPS and ultrasonic sensors.

imagery is a particular case of the SLAM problem, called bearing-only SLAM or boSLAM, in which bearing only sensors are used, a camera in this case. boSLAM is a partially observable problem [41], as the depth of the landmarks cannot be directly estimated. This entails a difficult landmark initialization problem which has been tackled with two basic approaches: delayed and un-delayed initialization. In the delayed initialization case, landmarks are not included in the SLAM system in the first observation, but when the angular baseline in between observations has grown large enough to ensure a good triangulation. This method has the advantage of using well conditioned landmarks, but the SLAM system cannot take advantage of the landmark until its localization is well conditioned. Several approaches have been proposed in this area such as [10] where a Particle Filter is used to initialize the landmark depth, or [11], where non-linear bundle adjustment over a set of observations is used to initialize the landmarks.

On the other hand, un-delayed approaches introduce the landmark in the SLAM system with the first observation, but some considerations have to be taken into account due to the fact that the landmarks are usually bad conditioned in depth, and then divergence problems may appear in the SLAM filter. Most existing approaches are based on multiple hypotheses, as in [20], where a Gaussian Mixture is used for landmark initialization in a Kalman Filter. Recent research [28] proposes the inverse depth parametrization in a single-hypothesis approach for landmark initialization.

The technique presented in this Section is based on a classical EKF that simultaneously estimates the pose of the robot (6 $df$) and a map of point features, as in [2, 3, 14, 24]. The main contributions is a new undelayed feature initialization that takes advantage of the scene normal plane estimation computed in the Homography-based odometry. Indeed, the technique cannot be considered as boSLAM because information from a range sensor is used, combined with the normal vector to the plane, to initialize the landmark depth.

The use of the estimated rotation and translation provided by the odometer as the main motion hypothesis in the prediction stage of the EKF is another contribution made by this approach. Complex non-linear models are normally used to estimate vehicle dynamics, due to the lack of odometers in UAVs. This leads to poor prediction hypotheses, in terms of accuracy, and then a significant reduction of the filter efficiency. In [19] a solution based on merging model-based estimation and inertial measurements from local sensors (IMUs) is proposed, resulting in an accuracy growth. The integration of the IMU is also considered here in order to improve the position estimation. Next paragraphs describe the structure and implementation of this filter.

### 3.1 The State Vector

The robot pose $\mathbf{p}_t$ is composed by the position and orientation of the vehicle at time $t$ in the World Frame (see Section 3.4), so:

$$\mathbf{p}_t = [\mathbf{t}_t, \mathbf{q}_t]^T = [x, y, z, q_x, q_y, q_z, q_w]^T \tag{18}$$

where $\mathbf{t}_t$ expresses the position at time $t$ of the UAV in the world coordinate frame, and $\mathbf{q}_t$ is the unitary quaternion that aligns the robot to the world reference frame at time $t$. Using quaternions increases (in one) the number of parameters for the orientation with respect to Euler angles, but simplifies the algebra and hence, the

error propagation. However, the quaternion normalization has to be taken into account after the prediction and update stages.

Landmarks will be represented by their 3D cartesian position in the World Frame $\mathbf{y}_n$. Thus, the state vector $\mathbf{x}_t$ is composed by the robot pose $\mathbf{p_t}$ and the set of current landmarks $\{\mathbf{y}_1, ..., \mathbf{y}_n\}$ so:

$$\mathbf{x}_t = \left[\mathbf{p}_t^T, \mathbf{y}_1^T, ..., \mathbf{y}_n^T\right]^T \tag{19}$$

### 3.2 Prediction Stage

Given the pose at time $t-1$, the odometer provides the translation with respect to the previous position (expressed in the $t-1$ frame) and the rotation that transforms the previous orientation into the new one (expressed in the $t$ frame). Taking into account the quaternions algebra, the state vector at time $t$ can be computed as:

$$\mathbf{t}_t = \mathbf{t}_{t-1} + \mathbf{q}_{t-1} \otimes \mathbf{t}_u \otimes \mathbf{q}_{t-1}^{-1} \tag{20}$$

$$\mathbf{q}_t = \mathbf{q}_u^{-1} \otimes \mathbf{q}_{t-1} \tag{21}$$

where $\mathbf{t}_u$ and $\mathbf{q}_u$ represent the estimated translation and rotation from the odometer, and $\otimes$ denotes quaternion multiplication. Notice that prediction does not affect the landmark position because they are assumed to be motionless.

Computing the odometry requires to carry out the image processing between consecutive images detailed in Section 2: feature tracking, homography estimation and, finally, odometry. The estimated translation and rotation covariance matrices are used to compute the process noise covariance matrix.

### 3.3 Updating Stage

From the whole set of features provided by the feature tracking algorithm used in the prediction stage, a small subset is selected to act as landmarks. The features associated to the landmarks are taken apart and not used for the homography estimation in order to eliminate correlations among prediction and updating. Thus, the number of landmarks must be a compromise between the performance of the EKF and the performance of the homography estimation (and thus, the odometry estimation). In addition, the computational requirements of the full SLAM approach has to be considered.

Experimental results allowed the authors to properly tune the number of landmarks and features used in the approach. A set of one hundred features are tracked from one image to another, and a subset of ten/fifteen well distributed and stable features are used as landmarks. Therefore, for each new image, the new position of this subset of features will be given by the feature tracking algorithm; this information will be used as measurement at time $t$, $\mathbf{z}_t$.

If the prediction stage was correct, the projection of each landmark into the camera would fit with the estimated position of the feature given by the tracking algorithm. If the landmark $\mathbf{y}_n$ corresponds to the image feature $\mathbf{m}_n = [u, v]$, following the camera projection model (Fig. 7):

$$\tilde{\mathbf{m}}_n = \mathbf{A}\left(\mathbf{q}_t^{-1} \otimes (\mathbf{y}_n - \mathbf{t}_t) \otimes \mathbf{q}_t\right) \tag{22}$$

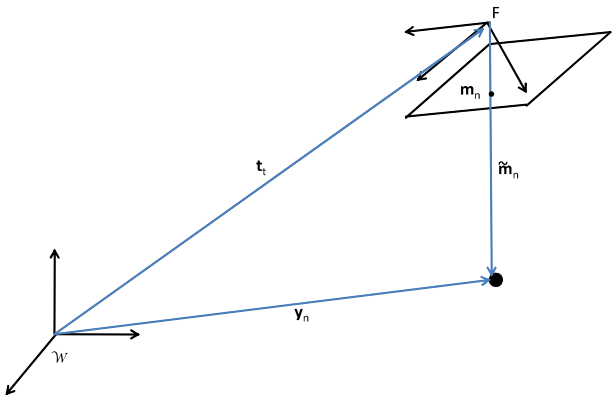


**Fig. 7** Projection of a landmark into the camera. The landmark is represented by a *black dot*, the translation of the camera focal ($F$) in the world frame ($\mathcal{W}$) is $\mathbf{t}_t$, the back-projection of the feature $\mathbf{m}_n$ is $\tilde{\mathbf{m}}_n$ and the position of the landmark in the world frame is $\mathbf{y}_n$

where $\mathbf{A}$ is the camera calibration matrix and $\tilde{\mathbf{m}}_n = [\tilde{u}, \tilde{v}, h]$, so the feature position is computed as $\mathbf{m}_n = [\tilde{u}/h, \tilde{v}/h]$. This measurement equation is applied to all the features correctly tracked from the previous image to the current one. The data association problem is solved by means of the feature matching algorithm.

In order to bound the computational cost needed for the SLAM approach, landmarks are not stored indefinitely in the EKF filter. Instead, they are maintained for a short period of time in the filter just to avoid transient occlusions, later they are automatically marginalized out from the filter and a new feature, provided by the tracker, initialized. If the corresponding landmarks are well conditioned, the measurement equation constraints the current position and orientation of the UAV.

### 3.4 Filter and Landmarks Initialization

The filter state vector will be initialized to a given position and orientation. This information can be provided by external devices such as GPS and IMU, and the process covariance matrix to the corresponding error information. The position can be also initialized to zero, so the first position is assumed as the origin and the corresponding covariances are zero too. This initial position defines the World Reference Frame where the landmarks and UAV pose are expressed.

In the following, a more sophisticated method for landmark initialization is proposed. When a new image feature is selected for being a landmark in the filter, it is necessary to compute its real position in the World frame. Due to the bearing only nature of the camera, the back-projection of the feature is given by a ray defined by the camera focal point and the image of the landmark. The proposed technique takes advantage of knowing the normal to the scene plane and the distance from the UAV to the ground at a given time. With this information the ground can be locally approximated by a plane and the landmark position as the intersection of the back-projection ray with this plane, as shown in Fig. 8.

If the World frame is aligned with the camera frame, the back-projection of the feature $\mathbf{m}_n = [u, v]$ will be the ray $\mathbf{r}$ defined by:

$$\mathbf{r} : \mathbf{A}^{-1}\tilde{\mathbf{m}}_n \tag{23}$$

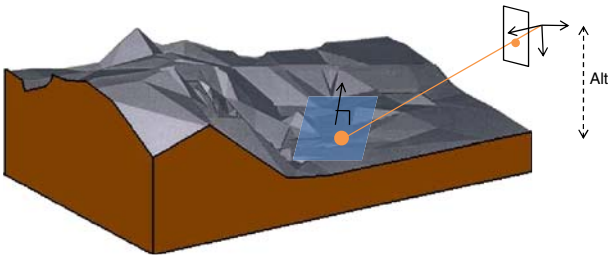


**Fig. 8** Landmark initialization representation

where $\mathbf{A}$ is the camera calibration matrix and $\tilde{\mathbf{m}}_n = [h\mathbf{m}_n, h]$. In addition, the odometer provides an estimation of the normal to the scene plane at time $t$ denoted as $\mathbf{n}_t$. Given the distance to the plane $d_t$, the plane $\Pi$ is defined as:

$$\Pi : d_t - \mathbf{n}_t^T \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{24}$$

Then, the landmark position will be computed as the intersection of the ray $\mathbf{r}$ with the plane $\Pi$. If Eqs. 23 and 24 are merged, the value of $\lambda$ can be easily computed as:

$$h = \left(\mathbf{n}_t^T \mathbf{A}^{-1} \tilde{\mathbf{m}}_n\right)^{-1} d_t \tag{25}$$

and the landmark can be computed as:

$$\mathbf{y}_n = \left(\mathbf{n}_t^T \mathbf{A}^{-1} \tilde{\mathbf{m}}_n\right)^{-1} d_t \mathbf{A}^{-1} \tilde{\mathbf{m}}_n \tag{26}$$

But this landmark is expressed in the camera coordinate frame. The UAV current position $\mathbf{d}_t$ and orientation $\mathbf{q}_t$ are finally used to express the landmark in the World frame:

$$\mathbf{y}_n = \mathbf{t}_t + \mathbf{q}_t \otimes \left(\left(\mathbf{n}_t^T \mathbf{A}^{-1} \tilde{\mathbf{m}}_n\right)^{-1} d_t \mathbf{A}^{-1} \tilde{\mathbf{m}}_n\right) \otimes \mathbf{q}_t^{-1} \tag{27}$$

There is a strong dependence of this approach on the planarity of the scene. The more planar the scene is, the better the plane approximation, leading to smaller noise in the plane normal estimation, and thus, to a better initialization.

Nevertheless, the back-projection procedure is still non-linear, and therefore, the Gaussian approximation for the errors has to be carefully considered. If the relative orientation of the ray $\mathbf{r}$ associated to a feature is near parallel with respect to the plane, the errors on the estimation can be high, and a Gaussian distribution will not approximate the error shape adequately. Then, only those landmarks for which the relative orientation of the ray and the plane is higher than 30 degrees will be considered in the initialization process.

### 3.5 Experimental Results on Homography-Based SLAM

To test the proposed approach, experiments with the HERO helicopter were carried out. The image sequence was gathered at 15 m of altitude with respect to the ground and with the camera pointed 45 degrees with respect to the helicopter horizontal.

It is important to remark that no close-loop was carried out during the experiment, although there are some loops present in the UAV trajectory, this subject is out of the

scope of this research work. Therefore, the result can be improved if a reliable data association algorithm is used for detecting and associating landmarks in the filter. The complete size of the trajectory is about 90 m long.

IMU information is used to express the results in the same frame than DGPS measurements. The results of the experiment are shown in Fig. 9, where the estimation

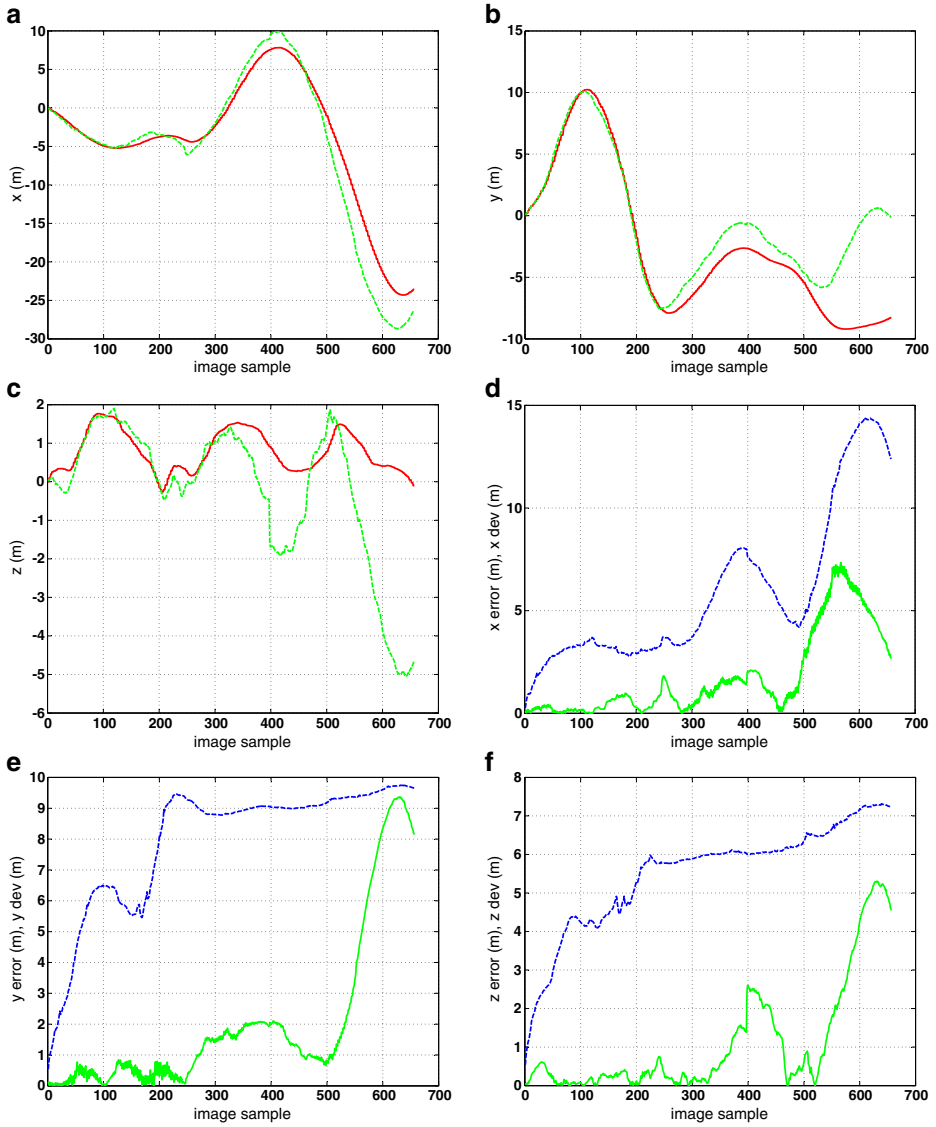


**Fig. 9** *Up* position estimation using the SLAM approach (*green dashed line*) and DGPS estimation (bi). *Down* error of the SLAM approach (*green solid line*) and estimated standard deviation (*blue dashed line*)

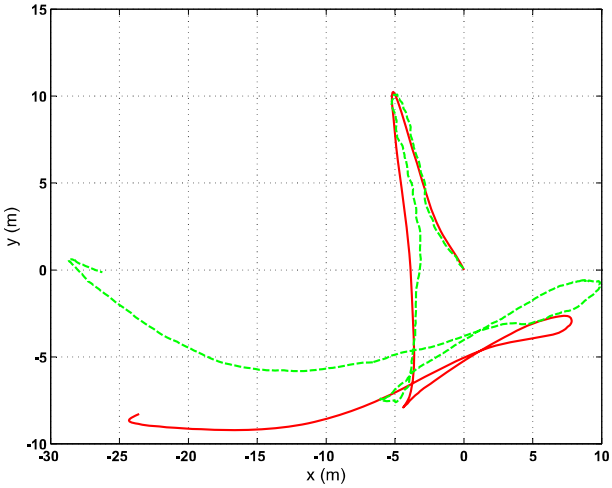


**Fig. 10** XY position estimation using the SLAM approach (*green dashed line*) and DGPS estimation (*red solid line*)

in each axis and the errors (with respect to DGPS outputs) are plotted. It can be seen how the uncertainty estimation is coherent with the measured errors. However, the position slowly diverges through time due to the absence of large loop closing. The instant orientation is not plotted because it is inherently taken into account in the computation of position. More details are shown in Fig. 10, where the XY DGPS trajectory is plotted together with the XY estimation.

### 3.6 Experimental Results Including an Inertial Measurement Unit

The errors shown in Fig. 9 are partially generated by a drift in the estimation of the UAV orientation. If the measurements of an inertial measurement unit (IMU) are incorporated into the SLAM approach, the errors introduced by the orientation estimation can be reset, and then the localization could be improved.

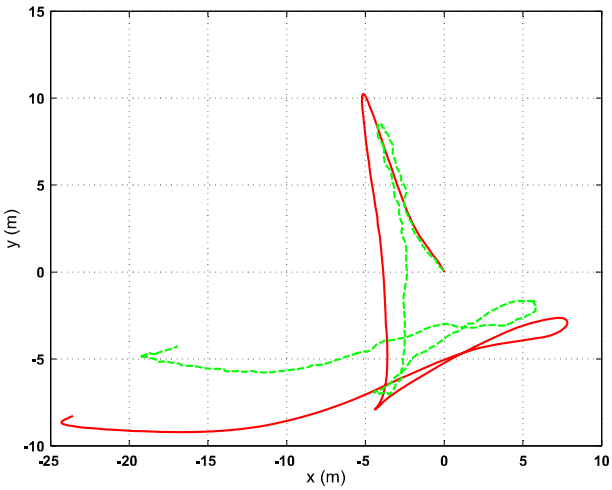


**Fig. 11** XY position estimation using the SLAM approach with IMU corrections (*green dashed line*) and DGPS estimation (*red solid line*)

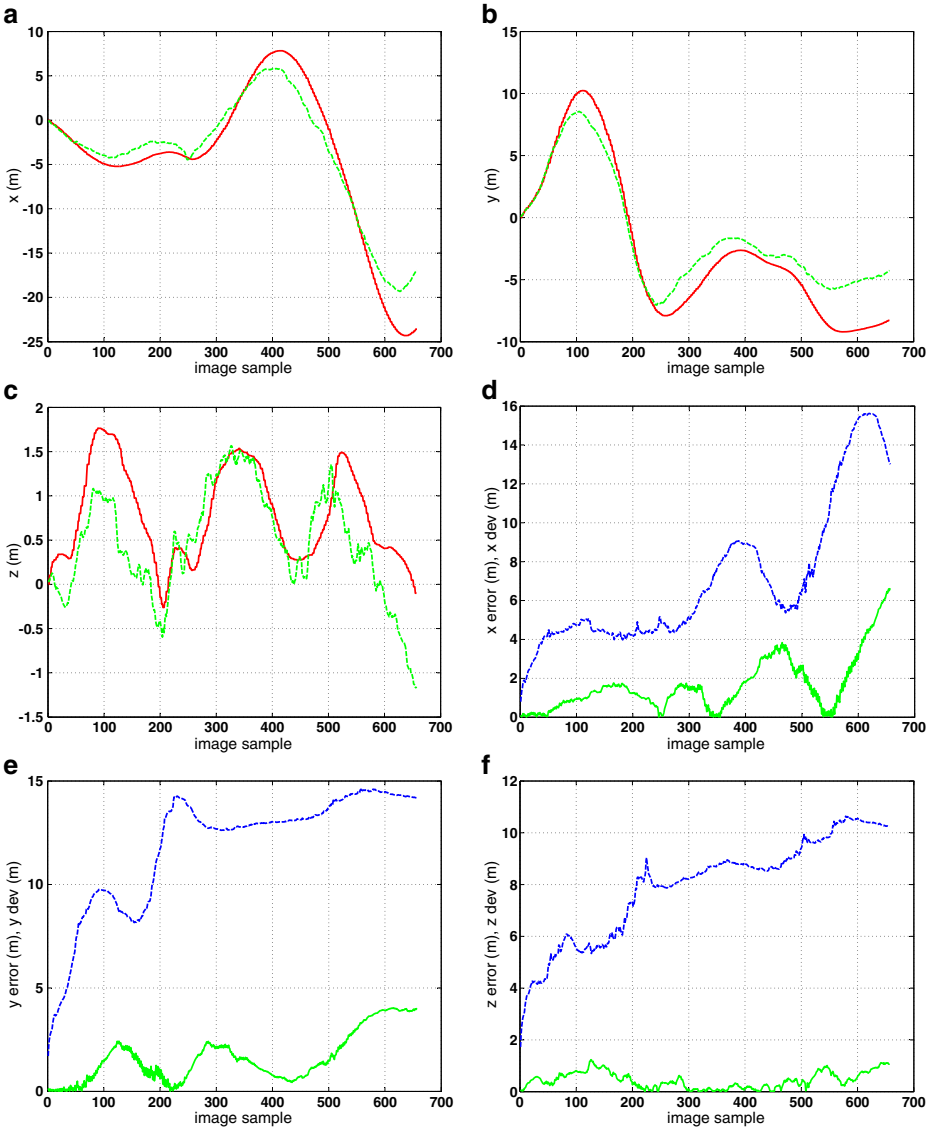


**Fig. 12** *Up* position estimation using the SLAM approach with IMU corrections (*green dashed line*) and DGPS estimation (*red solid line*). *Down* error of the SLAM approach with IMU corrections (*green solid line*) and estimated standard deviation (*blue dashed line*)

The proposed SLAM approach can be easily adapted to include the IMU information by integrating its measurement in the prediction stage of the EKF. The IMU provides the complete orientation, so there is no error integration and it is bounded by the accuracy of the device.

This approach has been tested with the same data set. The XY estimation is plotted in Fig. 11. Figure 12 shows the estimation compared to the DGPS measurement. It can be seen that the errors in Z and Y are significantly smaller while in X are slightly smaller with respect to the approach without considering the IMU.

## 4 Conclusions

The paper presents contributions to the vision-based navigation of aerial vehicles. It is proposed a visual odometry system for UAVs based on monocular imagery. Homographic models and homography decomposition are used to extract the real camera motion and the normal vector to the scene plane. A range sensor is used to obtain the scale factor of the motion. The paper shows the feasibility of the approach through experimental results with real UAVs.

An important aspect of the proposed odometry approach is the use of natural landmarks instead of beacons or visual references with known positions. A general-purpose feature tracking is used for this purpose. Although natural landmarks increase the applicability of the proposed techniques, they also increase the complexity of the problem to be solved. In fact, outlier rejection and robust homography estimation are required.

The paper also proposes a localization technique based on monocular vision. An Extended Kalman filter-based SLAM is successfully used to compute the localization and mapping. Two basic contributions to SLAM with UAVs are proposed. First, the use of a vision based odometry as main motion hypothesis for the prediction stage of the Kalman filter and, second, a new landmark initialization technique that exploits the benefits of estimating the normal to the scene plane. Both techniques are implemented and validated with a real UAV.

Although no large loops are closed in the experiments, the estimated position and covariance are coherent, so the result could be improved if a reliable data association algorithm is used for detecting and associating landmarks in the filter.

Future developments will consider different features with better invariance characteristics in order to close loops. It should be pointed out that the method can be applied to piece-wise planar scenes, like in urban scenarios

**Acknowledgements** The authors would like to thank the Technical University of Berlin for providing images to validate the homography-based odometry approach. In addition, the authors thank the support of Victor Vega, Fran Real and Ivan Maza during the experiments with HERO helicopter.

## References

1. Amidi, O., Kanade, T., Fujita, K.: A visual odometer for autonomous helicopter fight. In: Proceedings of the Fifth International Conference on Intelligent Autonomous Systems (IAS-5), June 1998
2. Betge-Brezetz, S., Hebert, P., Chatila, R., Devy, M.: Uncertain map making in natural environments. In: Proceedings of the IEEE International Conference on Robotics and Automation, vol. 2, pp. 1048–1053, April 1996
3. Betke, M., Gurvits, L.: Mobile robot localization using landmarks. IEEE Trans. Robot. Autom. **13**, 251–263 (1997)

4. Byrne, J., Cosgrove, M., Mehra, R.: Stereo based obstacle detection for an unmanned air vehicle. In: Proceedings 2006 IEEE International Conference on Robotics and Automation, pp. 2830–2835, May 2006
5. Caballero, F., Merino, L., Ferruz, J., Ollero, A.: A visual odometer without 3D reconstruction for aerial vehicles. applications to building inspection. In: Proceedings of the International Conference on Robotics and Automation, pp. 4684–4689. IEEE, April 2005
6. Caballero, F., Merino, L., Ferruz, J., Ollero, A.: Improving vision-based planar motion estimation for unmanned aerial vehicles through online mosaicing. In: Proceedings of the International Conference on Robotics and Automation, pp. 2860–2865. IEEE, May 2006
7. Caballero, F., Merino, L., Ferruz, J., Ollero, A.: Homography based Kalman filter for mosaic building. applications to UAV position estimation. In: IEEE International Conference on Robotics and Automation, pp. 2004–2009, April 2007
8. Conte, G., Doherty, P.: An integrated UAV navigation system based on aerial image matching. In: Proceedings of the IEEE Aerospace Conference, pp. 1–10 (2008)
9. Corke, P.I., Sikka, P., Roberts, J.M.: Height estimation for an autonomous helicopter. In: Proceedings of ISER, pp. 101–110 (2000)
10. Davison, A.: Real-time simultaneous localisation and mapping with a single camera. In: IEEE International Conference on Computer Vision, pp. 1403–1410, October 2003
11. Deans, M., Hebert, M.: Experimental comparison of techniques for localization and mapping using a bearings only sensor. In: Proceedings of the Seventh International Symposium on Experimental Robotics, December 2000
12. Demonceaux, C., Vasseur, P., Pegard, C.: Omnidirectional vision on UAV for attitude computation. In: Proceedings 2006 IEEE International Conference on Robotics and Automation, pp. 2842–2847, May 2006
13. Dickmanns, E.D., Schell, F.R.: Autonomous landing of airplanes using dynamic machine vision. In: Proc. of the IEEE Workshop Applications of Computer Vision, pp. 172–179, December 1992
14. Feder, H.J.S., Leonard, J.J., Smith, C.M.: Adaptive mobile robot navigation and mapping. Int. J. Rob. Res. **18**(7), 650–668 (1999) July
15. Garcia-Pardo, P.J., Sukhatme, G.S., Montgomery, J.F.: Towards vision-based safe landing for an autonomous helicopter. Robot. Auton. Syst. **38**(1), 19–29 (2001)
16. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press (2004)
17. Hrabar, S., Sukhatme, G.S.: Omnidirectional vision for an autonomous helicopter. In: Proceedings of the International Conference on Robotics and Automation, vol. 1, pp. 558–563 (2003)
18. Hygounenc, E., Jung, I.-K., Soueres, P., Lacroix, S.: The autonomous blimp project of LAAS-CNRS: achievements in flight control and terrain mapping. Int. J. Rob. Res. **23**(4-5), 473–511 (2004)
19. Kim, J., Sukkarieh, S.: Autonomous airborne navigation in unknown terrain environments. IEEE Trans. Aerosp. Electron. Syst. **40**(3), 1031–1045 (2004) July
20. Kwok, N.M., Dissanayake, G.: An efficient multiple hypothesis filter for bearing-only SLAM. In: Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 1, pp. 736–741, October 2004
21. Lacroix, S., Jung, I.K., Soueres, P., Hygounenc, E., Berry, J.P.: The autonomous blimp project of LAAS/CNRS - current status and research challenges. In: Proceeding of the International Conference on Intelligent Robots and Systems, IROS, Workshop WS6 Aerial Robotics, pp. 35–42. IEEE/RSJ (2002)
22. Langedaan, J., Rock, S.: Passive GPS-free navigation of small UAVs. In: Proceedings of the IEEE Aerospace Conference, pp. 1–9 (2005)
23. Lemaire, T., Lacroix, S., Solà, J.: A practical 3D bearing only SLAM algorithm. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2449–2454 (2005)
24. Leonard, J.J., Durrant-Whyte, H.F.: Simultaneous map building and localization for an autonomous mobile robot. In: Proceedings of the IEEE/RSJ International Workshop on Intelligent Robots and Systems, pp. 1442–1447, November 1991
25. Ling, L., Ridley, M., Kim, J.-H., Nettleton, E., Sukkarieh, S.: Six DoF decentralised SLAM. In: Proceedings of the Australasian Conference on Robotics and Automation (2003)
26. Mahony, R., Hamel, T.: Image-based visual servo control of aerial robotic systems using linear image features. IEEE Trans. Robot. **21**(2), 227–239 (2005)
27. Mejías, L., Saripalli, S., Campoy, P., Sukhatme, G.S.: Visual servoing of an autonomous helicopter in urban areas using feature tracking. J. Field. Robot. **23**(3–4), 185–199 (2006)

28. Montiel, J., Civera J, Davison, A.: Unified inverse depth parametrization for monocular SLAM. In: Robotics: Science and Systems, August 2006
29. Ollero, A., Ferruz, J., Caballero, F., Hurtado, S., Merino, L.: Motion compensation and object detection for autonomous helicopter visual navigation in the COMETS system. In: Proceedings of the International Conference on Robotics and Automation, ICRA, pp. 19–24. IEEE (2004)
30. Ollero, A., Merino, L.: Control and perception techniques for aerial robotics. Annu. Rev. Control, Elsevier (Francia), **28**, 167–178 (2004)
31. Papadopoulo, T., Lourakis, M.I.A.: Estimating the jacobian of the singular value decomposition: theory and applications. In: Proceedings of the 2000 European Conference on Computer Vision, vol. 1, pp. 554–570 (2000)
32. Proctor, A.A., Johnson, E.N., Apker, T.B.: Vision-only control and guidance for aircraft. J. Field. Robot. **23**(10), 863–890 (2006)
33. Remuss, V., Musial, M., Hommel, G.: Marvin - an autonomous flying robot-bases on mass market. In: International Conference on Intelligent Robots and Systems, IROS. Proceedings of the Workshop WS6 Aerial Robotics, pp. 23–28. IEEE/RSJ (2002)
34. Saripalli, S., Montgomery, J.F., Sukhatme, G.S.: Visually guided landing of an unmanned aerial vehicle. IEEE Trans. Robot. Autom. **19**(3), 371–380 (2003) June
35. Saripalli, S., Sukhatme, G.S.: Landing on a mobile target using an autonomous helicopter. In: Proceedings of the International Conference on Field and Service Robotics, FSR, July 2003
36. Shakernia, O., Vidal, R., Sharp, C., Ma, Y., Sastry, S.: Multiple view motion estimation and control for landing an aerial vehicle. In: Proceedings of the International Conference on Robotics and Automation, ICRA, vol. 3, pp. 2793–2798. IEEE, May 2002
37. Srinivasan, M.V., Zhang, S.W., Garrant, M.A.: Landing strategies in honeybees, and applications to UAVs. In: Springer Tracts in Advanced Robotics, pp. 373–384. Springer-Verlag, Berlin (2003)
38. Triggs, B.: Autocalibration from planar scenes. In: Proceedings of the 5th European Conference on Computer Vision, ECCV, vol. 1, pp. 89–105. Springer-Verlag, London, UK (1998)
39. Tsai, R.Y., Huang, T.S., Zhu, W.-L.: Estimating three-dimensional motion parameters of a rigid planar patch, ii: singular value decomposition. IEEE Trans. Acoust. Speech Signal Process. **30**(4), 525–534 (1982) August
40. Vidal, R., Sastry, S., Kim, J., Shakernia, O., Shim, D.: The Berkeley aerial robot project (BEAR). In: Proceeding of the International Conference on Intelligent Robots and Systems, IROS, pp. 1–10. IEEE/RSJ (2002)
41. Vidal-Calleja, T., Bryson, M., Sukkarieh, S., Sanfeliu, A., Andrade-Cetto, J.: On the observability of bearing-only SLAM. In: Proceedings of the 2007 IEEE International Conference on Robotics and Automation, pp. 1050–4729, April 2007
42. Volpe, J.A.: Vulnerability assessment of the transportation infrastructure relying on the global positioning system. Technical report, Office of the Assistant Secretary for Transportation Policy, August (2001)
43. Wu, A.D., Johnson, E.N., Proctor, A.A.: Vision-aided inertial navigation for flight control. In: Proc. of AIAA Guidance, Navigation, and Control Conference and Exhibit (2005)
44. Yakimenko, O.A., Kaminer, I.I., Lentz, W.J., Ghyzel, P.A.: Unmanned aircraft navigation for shipboard landing using infrared vision. IEEE Trans. Aerosp. Electron. Syst. **38**(4), 1181–1200 (2002) October
45. Zhang, Z., Hintz, K.J.: Evolving neural networks for video attitude and hight sensor. In: Proc. of the SPIE International Symposium on Aerospace/Defense Sensing and Control, vol. 2484, pp. 383–393 (1995) April