

# Motion compensation and object detection for autonomous helicopter visual navigation in the COMETS system\*

Aníbal Ollero, Joaquín Ferruz, Fernando Caballero,  
Sebastián Hurtado and Luis Merino

*Departamento de Ingeniería de Sistemas y Automática  
Escuela Superior de Ingenieros  
41010 Sevilla, Spain*

{aollero & ferruz}@cartuja.us.es

**Abstract** - This paper presents real time computer vision techniques for autonomous navigation and operation of unmanned aerial vehicles. The proposed techniques are based on image feature matching and projective methods. Particularly, the paper presents the application to helicopter motion compensation and object detection. These techniques have been implemented in the framework of the COMETS multi-UAV systems. Furthermore, the paper presents the application of the proposed techniques in a forest fire scenario in which the COMETS system will be demonstrated.

**Index Terms** - UAV; cooperative detection and monitoring; feature matching; image stabilization; homography.

## I. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) have increased significantly their flight performance and autonomous on-board processing capabilities in the last 10 years. These vehicles can be used in Field Robotics Applications where the ground vehicles have inherent limitations to access to the desired locations due to the characteristics of the terrain and the presence of obstacles that cannot be avoided. In these cases aerial vehicles may be the only way to approach the objective and to perform tasks such as data and image acquisition, localization of targets, tracking, map building, or even the deployment of instrumentation.

Particularly, unmanned helicopters are valuable for many applications due to their maneuverability. Furthermore, the hovering capability of the helicopter is very appreciated for event observation and inspection tasks. Many different helicopters with different degree of autonomy and functionalities have been presented in the last ten years (see for example [1], [2], [3], [4], [5]).

Most UAV autonomous navigation techniques are based on GPS and the fusion of GPS with INS information. However, computer vision is also useful to perceive the environment and to overcome GPS failures and accuracy degradation. Thus, the concept of visual odometer [1] was implemented in the CMU autonomous helicopter and this helicopter also demonstrated autonomous visual tracking capabilities of moving objects. Computer vision is also used for safe landing in [6]. In [7] a system for helicopter landing

on a slow moving target is presented. Vision based pose-estimation of unmanned helicopters relative to a landing target and vision-based landing of an aerial vehicle on a moving deck are also researched in [8] and [5]. Aerial image processing for an autonomous helicopter is also part of the WITAS project [9]. Reference [7] presents a technique for helicopter position estimation using a single CMOS camera pointing downwards, with a large field of view and a laser pointer to project a signature onto the surface below in such a way that it can be easily distinguished from other features on the ground. The perception system presented in [10] applies stereo vision, interest point matching and Kalman filtering techniques for motion and position estimation.

Motion estimation, object identification and geolocation by means of computer vision is also done in [11] and [12] in the framework of the COMETS project. In this paper new results of this project are also presented.

UAVs are increasingly used in many applications including surveillance and environment monitoring. Environmental disaster detection and monitoring is another promising application. Particularly, forest fire detection and monitoring are potential applications that attracted the attention of researchers and practitioners. In [13] the First Response Experiment (FiRE) demonstration of the ALTUS UAV (19,817 m altitude and 24 flight) for forest fire fighting is presented. The system is able to deliver geo-rectified image file within 15 minutes of acquisition. In [11] and [12] the application of computer vision techniques and UAVs for fire monitoring using aerial images is proposed. The proposed system provides in real time the coordinates of the fire front by means of geo-location techniques. Instead of expensive high performance UAVs, the approach is to use multiple low cost aerial system. This paper also presents experiments in a forest fire scenario using this approach in the framework of the COMETS multi-UAV project funded by the European Commission under the IST program.

In the next section the COMETS system is introduced. Then, a feature matching method based on previous work of some of the authors is summarized. This method is applied in the next two sections to motion compensation, required for object detection. Then, experimental results are described. Finally, the conclusions and references are presented.

---

\* This research work has been partly supported by the COMETS (IST-2001-34304) and CROMAT (DPI2002-04401-C03-03) projects.

## II. THE COMETS SYSTEM

This research work has been developed in the framework of the COMETS project. The main objective of COMETS is to design and implement a distributed control system for cooperative detection and monitoring using heterogeneous Unmanned Aerial Vehicles (UAVs). Distributed sensing techniques which involve real-time processing of aerial images play an important role. Although the architecture is expected to be useful in a wide spectrum of environments, COMETS will be demonstrated in a fire fighting scenario.

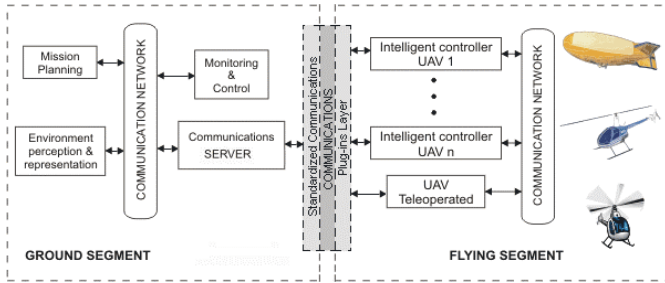


Fig. 1 Architecture of the COMETS system

COMETS (see Fig. 1) includes heterogeneous systems both in terms of vehicles (helicopters [3] and airships [10] have been currently integrated) and on board processing capabilities ranging from fully autonomous aerial systems to conventional radio controlled systems. The perception functionalities of the COMETS system can be implemented on-board the vehicles or on ground stations, where low cost and light aerial vehicles without enough on-board processing capabilities are used. A system like this poses significant difficulties on image processing activities. Being a distributed, wireless system with non-stationary nodes, bandwidth is a non-negligible limit. In addition, small aerial vehicles impose severe limits on the weight, power consumption and size of the on board computers, making necessary to run most of the processing off-board. The same constraints and its high cost rule out high-performance gimbals which are able to cancel vibration of on-board cameras. Thus, image processing should be able to extract useful information on low frame rate (one to two frames per second), compressed video streams where camera motion is largely uncompensated. A precondition for many detection and monitoring algorithms is electronic image stabilization, which in turn depends on a sufficiently reliable and robust image matching method, able to handle the high and irregular apparent motion that is frequently found in aerial uncompensated video, even when the platform is a hovering helicopter. This function will be considered in section III.

The perception system in COMETS consists of the Application Independent Image Processing (AIIP) subsystem, the Detection Alarm Confirmation and Localization (DACLE) subsystem, and the Event Monitoring System (EMS). This paper describes several functions of the AIIP subsystem. These functions are used by DACLE and EMS. Particularly,

image stabilization and object detection are described in the following and implemented in the helicopter shown in Fig. 3 developed jointly by the University of Seville and the Helivision company. Object detection can be useful in a multi-UAV system for security reasons; emergency collision avoidance would need to detect nearby crafts. On the other hand, surveillance activities would also need some way to detect and track mobile objects on the ground, such as cars.



Fig. 2 University of Seville-Helivision helicopter flying in experiments of the COMETS project (May 2003).

## III. FEATURE MATCHING METHOD

### A. Relations to previous work

The computation of the approximate ground plane homography needs a number of good matching points between pairs of images in order to work robustly. The image matching method used in this work is related to the described in [14], although significant improvements have since been made.

In [14] corner points were selected using the criteria described in [15]; each point was the center of a fixed-size window which is used as template in order to build matching window sequences over the stream of video images. Window selection provides for initial startup of window sequences as well as candidates (called *direct* candidates) for correlation-based matching tries with the last known template window of a sequence. The selection of local maxima of the corner detector function assured stable features, so window candidates in any given image were usually near the right matching position for some window sequence.

The correlation-based matching process with direct candidates within a search zone allowed to generate a matching pair data base, which described possibly multiple and incompatible associations between tracked sequences and candidates. A disambiguation process selected the right window to window matching pairs by using two different constraints: least residual correlation error and similarity between clusters of features.

The similarity of shape between regions of different images is verified by searching for clusters of windows whose members keep the same relative position, after a scale factor is applied. For a cluster of window sequences,

$\Gamma = \{\Phi_1, \Phi_2 \dots \Phi_n\}$ , this constraint is given by the following expression:

$$\left| \frac{\|w_{\Phi_k}, w_{\Phi_l}\|}{\|v_{\Phi_k}, v_{\Phi_l}\|} - \frac{\|w_{\Phi_p}, w_{\Phi_q}\|}{\|v_{\Phi_p}, v_{\Phi_q}\|} \right| \leq k_p \quad \forall \Phi_k, \Phi_l, \Phi_p, \Phi_q \in \Gamma \quad (1)$$

In (1)  $k_p$  is a tolerance factor,  $w_i$  are candidate windows in the next image and  $v_i$  are template windows from the preceding image. The constraint is equivalent to verify that the euclidean distances between windows in both images are related by a similar scale factor; thus, the ideal cluster would be obtained when euclidean transformation and scaling can account for the changes in window distribution.

Cluster size is used as a measure of local shape similarity; a minimum size is required to define a valid cluster. If a matching pair cannot be included in at least one valid cluster, it will be rejected, regardless of its residual error.

### B. New strategy for feature matching

The new approach uses the same feature selection procedure, but its matching strategy is significantly different.

First, the approach ceases to focus in individual features. Now clusters are not only built for validation purposes; they are persistent structures which are expected to remain stable for a number of frames, and are searched for as a whole.

Second, the disambiguation algorithm changes from a relaxation procedure to a more efficient predictive approach, similar to the one used in [16] for contour matching. Rather than generating an exhaustive data base of potential matching pairs as in [15], only selected hypothesis are considered. Each hypothesis, with the help of the persistent cluster data base, allows to define reduced search zones for sequences known to belong to the same cluster as the hypothesis, if a model for motion and deformation of clusters is known. Currently the same approach expressed in (1) is kept, refined with an additional constraint over the maximum difference of rotation angle between pair of windows:

$$\left| \alpha_{\Phi_k \Phi_l} - \alpha_{\Phi_p \Phi_q} \right| \leq \gamma_p \quad \forall \Phi_k, \Phi_l, \Phi_p, \Phi_q \in \Gamma \quad (2)$$

Where  $\alpha_{rs}$  is the rotation angle of the vector that links windows from sequences  $r$  and  $s$ , if the matching hypothesis is accepted, and  $\gamma_p$  is a tolerance factor. Although the cluster model is adequately simple and seems to fit the current applications, more realistic local models such as affine or full homography could be integrated in the scheme without much difficulty.

It is easy to verify that two hypothesized matching pairs allow to predict the position of the other members of the cluster, if their motion can be modelled approximately by euclidean motion plus scaling. Using this model, the generation of *candidate clusters* for a previously known cluster can start from a primary hypothesis, namely the

matching window for one of its window sequences (see Fig. (3)). This assumption allows to restrict the search zone for other sequences of the cluster, which are used to generate at least one secondary hypothesis. Given both hypothesis, the full structure of the cluster can be predicted with the small uncertainty imposed by the tolerance parameters  $k_p$  and  $\alpha_p$ , and one or several candidate clusters can be added to a data base. The creation of any given candidate cluster can trigger the creation of others for neighbour clusters, provided that there is some overlap among them; in Fig. (1), for example, the creation of a candidate for cluster 1 can be used immediately to propagate hypothesis and find a candidate for cluster 2. Direct search of matching windows is thus kept to a minimum.

At the final stage of the method, the best cluster candidates are used to generate clusters in the last image, and determine the matching windows for each sequence.

The practical result of the approach is to drastically reduce the number of matching tries, which are by far the main component of processing time when a great number of features have to be tracked, and large search zones are needed to account for high speed image plane motion. This is the case in non-stabilized aerial images, specially if only relatively low frame rate video streams are available.

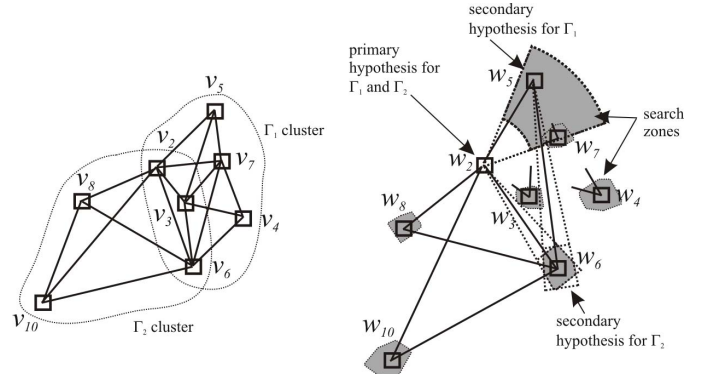


Fig. 3 Generation of cluster candidates

### C. Other features

In addition to the cluster based, hypothesis driven approach, other improvements have been introduced in the matching method.

- Temporary loss of sequences is tolerated through the prediction of the current window position computed with the known position of windows that belong to the same cluster; this feature allows to deal with sporadic occlusion or image noise.
- Normalized correlation is used instead of the sum of squared differences (SSD) used in [14], in order to achieve greater immunity to change in lighting conditions. The higher computational cost has been reduced with more efficient algorithms that involve applying the method described in [15] between a previously normalized template and candidate windows.

#### IV. MOTION COMPENSATION AND OBJECT DETECTION

Motion compensation can be achieved for specific configurations through the computation of homography between pairs of images.

##### A. Homography computation

If a set of points in the scene lies in a plane, and they are imaged from two viewpoints, then the corresponding points in images  $i$  and  $j$  are related by a plane-to-plane projectivity or planar homography [17],  $H$ :

$$s\tilde{m}_i = H\tilde{m}_j \quad (3)$$

where  $\tilde{m}_k = [u_k, v_k, 1]$  is the vector of homogenous image coordinates for a point in image  $k$ ,  $H$  is a  $3 \times 3$  non-singular matrix and  $s$  is a scale factor. The same equation holds if the image to image camera motion is a pure rotation. Even though the hypothesis of planar surface or pure rotation may seem too restrictive, they have proved to be frequently valid for aerial images. An approximate planar surface model usually holds if the UAV flies at a sufficiently high altitude, while an approximate pure rotation model holds for a hovering helicopter. Thus, the computation of  $H$  will allow under such circumstances to compensate for camera motion.

Since  $H$  has only eight degrees of freedom, we only need four correspondences to determine  $H$  linearly. In practice, more than four correspondences are available, and the overdetermination is used to improve accuracy. For a robust recovery of  $H$ , it is necessary to reject outlier data. In the proposed application, outliers will not always be wrong matching pairs; image zones where the homography model will not hold (moving objects, buildings or structures which break the planar hypothesis) will also be regarded as outliers, although they may offer potentially useful information. The overall design of the outlier rejection procedure used in this work, is based on LMedS (Least Median Square Stimator) and further refined by the Fair M-estimator [18], [19], [20], [21].

##### B. Optimized motion compensation algorithm

Once the homography matrix  $H$  has been computed, it is possible to compute from (3) the position in image  $j$ ,  $\tilde{m}_j = [u_j, v_j, 1]$ , where the point in image  $i$ ,  $\tilde{m}_i = [u_i, v_i, 1]$ , has moved. As  $u_j, v_j$  are in general non-integer coordinates, some interpolation algorithm such as bilinear or nearest-neighbour will have to be used to obtain the motion compensated image  $j$ .

As the COMETS system needs to operate in real-time, it was necessary to optimize the motion compensation process, which was intended to support other higher level processing. As the computation of  $u_j, v_j$  was found to spend a significant portion of the processing time devoted to motion compensation, an approximate optimized method has been designed.

If the straightforward computation is used, each coordinate pair needs at least 14 floating point arithmetic

operations, two of them divisions, which are usually significantly slower than multiplications or additions:

$$\begin{aligned} s(u_i, v_i) &= u_i H_{31} + v_i H_{32} + H_{33} \\ u_j &= (u_i H_{11} + v_i H_{12} + H_{13}) / s(u_i, v_i) \\ v_j &= (u_i H_{21} + v_i H_{22} + H_{23}) / s(u_i, v_i) \end{aligned} \quad (4)$$

Under an affine transformation,  $s(u_i, v_i) = H_{33}$ ; if  $H$  is normalized to set  $H_{33} = 1$ , the number of operations per pixel would be reduced to 8: Four additions, four multiplications and no divisions. For a general homography matrix, a linear approximation can be used:

$$\begin{aligned} u_j &\approx au_i + b \\ v_j &\approx cv_i + d \end{aligned} \quad (5)$$

Where the coefficients  $a, b, c, d$  are computed for each row of the image; better results are obtained if the nonlinear transformation is stepwise linearized by computing  $a, b, c, d$  in a number of intervals which depends on the nonlinearity of the specific transformation. As (4) shows, nonlinearity is linked to the function  $s(u_i, v_i)$ , and will decrease with the range of variation of  $s_{nl} = s(u_i, v_i) - H_{33} = u_i H_{31} + v_i H_{32}$ . As the preceding expression defines a plane over the pixel coordinate space, the maximum absolute value of  $s_{nl}$ ,  $|s_{nl}|_M$ , will be reached on the corners of the image, and can be easily computed. The following heuristic expression is used to determine the optimal number of linearization intervals:

$$n_s = \text{ceil}(5.68 |s_{nl}|_M^{0.5024}) \quad (6)$$

As a result of this optimization, the computation time for motion compensation in images of  $384 \times 287$  pixels decreases from 80 to 30 ms. in a Pentium 3 at 1Ghz. The combined execution of JPEG decompressing, feature matching and motion compensation allows to deal with displacements of up to 100 pixels at a rate of about three frames per second.

##### C. Object detection

Object detection is an example of processing that can be performed with the motion compensated stream of images. Moving objects can be detected by first segmenting regions whose motion is independent from the ground reference plane. Object detection can be further refined by searching for specific features in such regions.

Independent motion regions are detected by processing the outliers detected during the computation of  $H$ . Points where  $H$  cannot describe motion may appear not only because of errors in the matching stage; the local violation of the planar assumption, or the presence of mobile objects will generate outliers as well. In a second stage, a specific object can be identified among the candidate regions generated by the independent motion detection procedure. Temporal consistency constraints can be used for this purpose, as well as

known features of the specific object of interest. In the current approach, color signature is used to identify nearby aircrafts, as shown in section V.

## V. EXPERIMENTAL RESULTS

Figure 4 shows the results of tracking on a pair of images; pictures on top show the tracked windows, while pictures on the bottom show a magnified detail with a cluster. For clarity, only successfully tracked or newly selected windows are displayed; window 103, circled in white in the lower left picture, is temporarily lost because it moves behind the black overlay, but would be predicted from the known position of the other members.

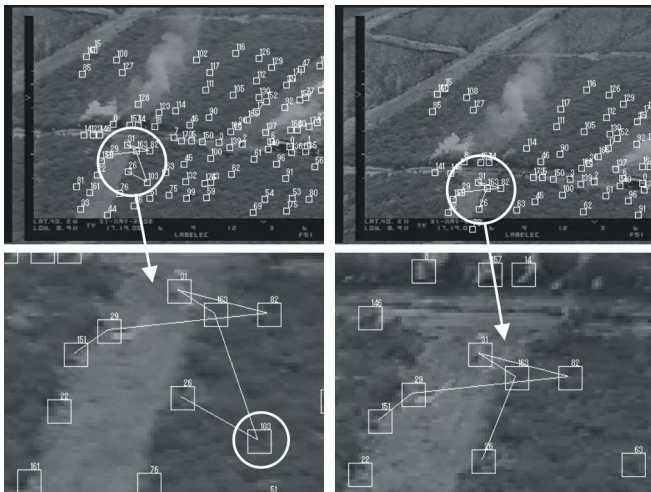


Fig. 4 Feature matching method.

In Fig. 5, the image matching and motion compensation algorithms are run over a non-compensated stream of images. The upper pictures are original images; below are their compensated versions, which show the changing features (fire and smoke) over a static background. In this case only the common field of view is represented, while the rest is clipped; the black zone on the right of the second image is beyond its limits. Long compensated sequences can be visualized in the official web page of the COMETS project, <http://www.comets-uavs.org>.

In Fig. 6 the outliers shown in the upper right image, are detected among the tracked windows in the upper left picture when homography is computed. The outlier points are clustered to identify areas that could belong to the same object, or discarded, if they are too sparse. This is shown in the lower image of Fig. 5, where three clusters are created, marked with a white square; only one of them is selected, because its color signature is the expected for a helicopter, different from other mobile objects like fire. In Fig. 7, a conventional helicopter is identified and tracked during four consecutive frames by using the described approach.

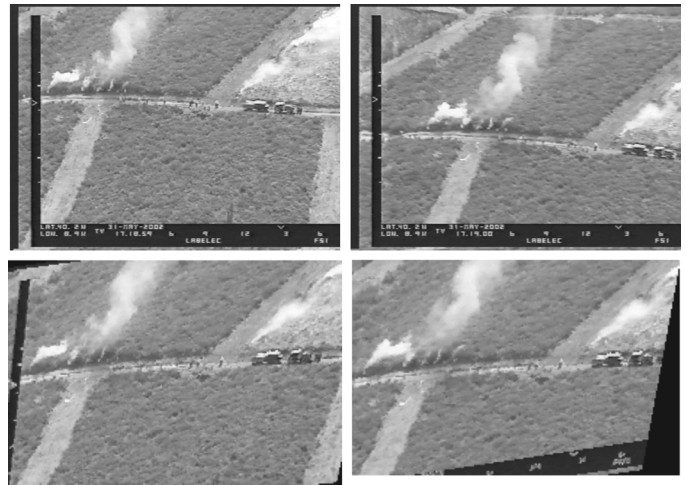


Fig. 5 Motion compensation results.

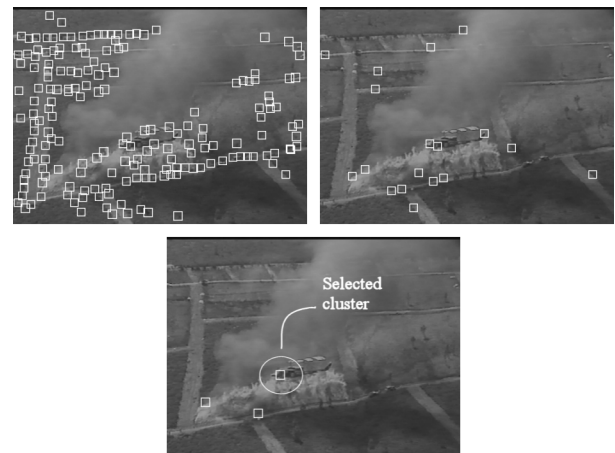


Fig. 6 Independent motion detection through outlier analysis.



Fig. 7 Tracking of a nearby helicopter.

## VI. CONCLUSIONS

In this paper some significant results of vision-based object detection and UAV motion compensation results have

been presented. They work on non-stabilized image streams captured from low-cost unmanned flying platforms, in the framework of the real-time distributed COMETS system.

#### ACKNOWLEDGMENT

The authors wish to thank the collaboration of the other COMETS partners, specially Miguel Ángel González, from Helivision, whose expertise has been helpful to obtain the experimental results.

#### REFERENCES

- [1] O. Amidi, T. Kanade and K. Fujita, "A visual odometer for autonomous helicopter flight," Proceedings of IAS-5, 1998.
- [2] A.H. Fagg, M.A., Lewis, J.F. Montgomery and G.A. Bekey, "The USC Autonomous Flying vehicle: An experiment in real-time behaviour-based control," Proceedings of the 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp 1173-1180, 1993.
- [3] V. Remuß, M. Musial and G. Hommel, "MARVIN – An Autonomous Flying Robot-Bases On Mass Market," I2002 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems – IROS 2002, Proc. Workshop WS6 Aerial Robotics, pp 23-28. Lausanne, Switzerland, 2002.
- [4] M. Sugeno, M.F. Griffin, and A. Bastia, "Fuzzy Hierarchical Control of an Unmanned Helicopter", 17<sup>th</sup> IFSA World Congress, pp. 179-182.
- [5] R. Vidal, S. Sastry, J. Kim, O. Shakernia and D. Shim, "The Berkeley Aerial Robot Project (BEAR)," 2002 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems – IROS 2002. Proc. Workshop WS6 Aerial Robotics, pp 1-10. Lausanne, Switzerland, 2002.
- [6] P.J. Garcia-Pardo, G.S. Sukhatme and J.F. Montgomery, "Towards Vision-Based Safe Landing for an Autonomous Helicopter," Robotics and Autonomous Systems, Vol. 38, No. 1, pp. 19-29, 2001.
- [7] S. Saripally, and G.S. Sukhatme, "Landing on a mobile target using an autonomous helicopter," IEEE Conference on Robotics and Automation, 2003.
- [8] O. Shakernia, O., R. Vidal, C. Sharp, Y. Ma and S. Sastry, "Multiple view motion estimation and control for landing and unmanned aerial vehicle," Proceedings of the IEEE International Conference on Robotics and Automation, 2002.
- [9] K. Nordberg, P. Doherty, G. Farneback, P-E. Forssen, G. Granlund, A. Moe and J. Wiklund, "Vision for a UAV helicopter," 2002 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems–IROS 2002. Proc. Workshop WS6 Aerial Robotics, pp 29-34. Lausanne, 2002.
- [10] S. Lacroix, I-K. Jung, P. Soueres, E. Hygounenc and J-P. Berry, "The autonomous blimp project of LAAS/CNRS - Current status and research challenges," 2002 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems – IROS 2002. Proc. Workshop WS6 Aerial Robotics, pp 35-42. Lausanne, Switzerland, 2002.
- [11] L. Merino and A. Ollero, "Forest fire perception using aerial images in the COMETS Project," 2002 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems – IROS 2002. Proc. Workshop WS6 Aerial Robotics, pp 11-22. Lausanne, Switzerland, 2002.
- [12] L. Merino, and A. Ollero, "Computer vision techniques for fire monitoring using aerial images," Proc. of the IEEE Conference on Industrial Electronics, Control and Instrumentation IECON 02 Seville (Spain), 2002.
- [13] V. G. Ambrosia, "Remotely piloted vehicles as FIRE imaging platforms: The future is here!," Wildfire, pp. 9-16, 2002.
- [14] J. Ferruz and A. Ollero, "Real-time feature matching in image sequences for non-structured environments. Applications to vehicle guidance," Journal of Intelligent and Robotic Systems, Vol. 28, pp. 85-123, June 2000.
- [15] C. Tomasi, "Shape and motion from image streams: a factorization method (*Ph.D. Thesis*)," Carnegie Mellon University, 1991.
- [16] N. Ayache, "Artificial Vision for Mobile Robots," The MIT Press, Cambridge, Massachusetts, 1991.
- [17] J. Semple and G. Kneebone, "Algebraic projective geometry," Oxford University press, 1952.
- [18] G. Xu and Z. Zhang, "Epipolar geometry in stereo, motion and object recognition," Kluwer Academic Publishers, 1996.
- [19] Z. Zhang, "A new multistage approach to motion and structure estimation: from essential parameters to Euclidean motion via fundamental matrix," RR-2910, Inria, 1996.
- [20] Z. Zhang, "Parameters estimation techniques. A tutorial with application to conic fitting," RR-2676, Inria, October 1995.
- [21] O. Faugeras, Q. Luong and T. Papadopoulos, "The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications", MIT Press, 2001.