Issues in cooperative air/ground robotic systems

Simon Lacroix[1] and Guy Le Besnerais[2]

[1] LAAS/CNRS
University of Toulouse
7, Ave du Colonel Roche
31077 Toulouse Cedex 4, France
Simon.Lacroix@laas.fr

[2] Onera/DTIM
29 avenue de la Division Leclerc
BP 72, 92322 CHATILLON Cedex, France
Guy.Le_Besnerais@onera.fr

Summary. The possibility to jointly deploy aerial and ground robots makes sense in many application contexts of field robotics. There are indeed several cooperation schemes in which the complementarity of such heterogeneous robots can be exploited to enhance the efficiency of autonomous robotic operations.

This paper analyses the problems raised by the cooperation of ground and aerial robots. Besides the usual issues brought forth by the integration of cooperating heterogeneous autonomous systems, such systems raise particularly challenging problems for environment perception and modelling. On the basis of a review of the state of the art in the area, the paper focuses on this particular issue. It analyzes the required developments to tackle it, and sets forth some working directions.

1 Introduction

The robotics community is paying more and more attention to the development of aerial robots. As opposed to drones that execute pre-programmed mission, aerial robots exhibit decisional autonomy: they are meant to achieve high level missions, such as mapping or monitoring a given area, with little human interactions. Aerial robots rise a wide number of robotics research areas, that ranges from the study of innovative flying concepts and the associated flight control algorithms (especially for micro drones) to high level mission planning and execution control, via real time environment mapping. Such robots are naturally aimed for exploration and monitoring missions in remote or hostile areas, as they can easily gather detailed information on the environment, without exposing themselves or the operators to any danger.

But the information gathered by aerial means are not always complete, especially in cluttered urban areas, and aerial robots can hardly physically intervene on the environment. Actually, in all the applications contexts where the development of exploration and intervention robotics is considered, air/ground robotic ensembles bring forth several opportunities from an operational point of view. Be it for environment monitoring and surveillance, demining or reconnaissance, a variety of scenarios can be envisaged in which the complementarity of both kinds of systems can yield more autonomous, efficient and robust operations. For instance, aircrafts can operate in a preliminary phase, in order to gather information that will later be used in both mission planning and execution for the ground vehicles. But one can also foresee cooperative scenarios, where aircrafts would support ground robots with communication links and global information, or where both kinds of machines would cooperate to fulfill a given mission.

After a brief review of the related work, the first section synthesizes various cooperation schemes for air/ground robotics systems, and outlines the associated main research issues. Besides usual decisional issues for multi-robot heterogeneous systems, it appears that one of the most important problem to solve is the ability to fuse aerial and ground perception data within integrated environment representations. Section 3 is dedicated to this issue: it sets forth the essential functions to achieve, and discusses possible solutions.

2 Air/Ground cooperation

2.1 Related work

The literature does not abound with contributions on the cooperation between aerial and terrestrial robots. Nevertheless, there have been these recent years some significant achievements, that exhibit some particular instances of cooperation.

The experiments conducted at the GRASP Laboratory [CGK+04] focus on the localization of a ground robot with the help of images gathered by a blimp. The blimp positions are estimated by matching ground features, and estimates of the UGV positions are obtained by detecting it in the blimp images. Further developments using a Decentralized Data Filter and the construction of radio-connectivity maps have been presented in [CCG+05]. The Decentralized Data Filter developed at ACFR [RNSDW02] has been applied in [GBK+04] to locate static ground targets, in a case where both kinds of vehicles are located by GPS and perceive the target with vision. Note that in these contributions, the experiments are set up so as to ease the sensor processing (targets are known patterns, located on a horizontal plane and easily detected an associated in the captured images).

Other cooperation schemes in which the UGV exploits data gathered by a UAV have been demonstrated. In [SKHR02], the UAV acts as a "flying eye" scouting for the UGV: it gathers traversability information that are fused with the UGV data for path planning purposes. More impressive results are presented in [VDH06], in which the UAV builds a 3D map that is also used for

UGV localization purposes, using "spin-images" to register the ground and aerial 3D maps [Joh00]. Substantial enhancements have been developed upon the spin-images approach, so as to be able to find feature matches between the maps independently built by the UAV and the UGV. 3D information is essential in these contributions, and is obtained thanks to stereovision of Lidars for both kinds of vehicles.

Other ambitious scenarios that involve aerial and ground robots in a coherent system are envisaged in [MTW$^+$06, HBB$^+$05].

2.2 Cooperation scenarios

The operational functions that can benefit from field robotics systems consists in exploration, monitoring or intervention tasks. These basic tasks can be declined and assembled in various mission scenarios, depending on the application context. For instance, "Search and rescue" missions calls for both an exploration phase (finding the position of victims) and an intervention phase (reaching the victims to provide them with assistance).

If these tasks can be achieved by a single robot, be it terrestrial or aerial, they can obviously be more efficiently achieved by an ensemble of aerial and ground robots, that exhibits a wider spectrum of complementary capacities, from both a perception and motion point of view. For short, aerial vehicles can move rapidly for one place to the other, they can provide both global and precise views on the environment, and are less prone to communication and GPS signals outages than ground robots. On the contrary, ground robots can carry larger payloadsf and endure longer missions. They operate more closely to the environment, and provide the capacities to intervene on the environment, to deploy sensing and communicating devices, and to perceive details hidden to the UAVs – *e.g.* under the canopy or on building facades.

As in any multi-robot system, a variety of integrating schemes can be foreseen. But the complementarities of aerial and ground robots drive the definition of specific cooperation schemes, that one can cast in the following three types:

- *Aerial robots assist ground robots.* Aerial robots can provide the ground robots with information related to the environment (traversability maps, landmark maps). They can also straightforwardly localize the ground robots by perceiving them, or provide communication links with the remote operator station and between ground robots.
- *Ground robots assist aerial robots.* Ground robots can assist the aerial robots by detecting cleared landing areas, or by providing them energy or the possibility to transport or recover them (as considered in [SKHR02, pro05] for instance).
- *Ground an aerial robots cooperate to achieve a task.* Exploration and surveillance tasks can obviously be jointly achieved by teams of aerial and ground robots. Similarly, target detection and tracking tasks can benefit by the enhanced observation capacities brought by such teams – and more

generally any combination of the two assistance schemes mentioned above can be foreseen.

2.3 Issues raised by air/ground cooperating systems

As for any multi-robot system, the deployment of aerial and ground robots calls for the resolution of problems related to perception, decision and action. Besides the fact that the robots are heterogeneous, provided their perception and action capacities are properly modelled, the decision issues do not exhibit any specificity different from other multi robot systems (task allocation, joint tasks planning, coordinated execution control... all these functions being embedded in a decentralized architecture). Action issues might call for specific developments to ensure particular tasks – such as servoing the UAV motions to the UGV motions, but do not raise any new or unsolved problem.

There is no doubt that it is in the area of perception that most of the difficulties arise, as the data gathered by aerial and ground means have very different characteristics. Research in the integration of such data is all the more important for air/ground robotics systems that the ability to build and share environment representations is a required basis to establish cooperation schemes, and is thus required by any of the tasks mentioned in section 2.2 – even for missions whose objectives are not specified in perception terms.

As an illustration, let's consider a simple scenario similar to the "flying eye" approach studied in [SKHR02], in which the aerial robot provides a ground robot with traversability information, the mission being defined as a distant goal to reach for the ground robot. A traversability map expresses the *cost* of traversing given areas for a UGV, or more simply consists in a binary description of obstacle/free areas: Fig. 1 shows such a map built from a monocular aerial image sequence, and Fig. 2 shows a traversability map built by a UGV from a sequence of stereoscopic images.

In such a scenario, the traversability map built by the UAV must be *fused* with the one built by the UGV: the UAV can indeed not assess the traversability under the tree canopy for instance, or can not detect small obstacles over which the UGV can not go through. Fusing these maps require that they are precisely spatially registered: in the absence of any precise localization mean for both robots, a realistic assumption in most cases, and especially in military applications where GPS localization can not be taken for granted, the only solution to register the two maps is to rely on *landmarks* that are *detected and matched by the two kinds of robots*[3].

This simple example scenario shows that the ability to integrate aerial and ground data within *various dedicated environment representations* is an essential prerequisite to have both kinds of robots cooperate to achieve a given mission.

[3] Note also that both maps must be spatially consistent beforehand: for that purpose both robots can exploit landmarks they are able to rely upon to maintain precise position estimates – these landmarks being not necessarily shared between the robots.

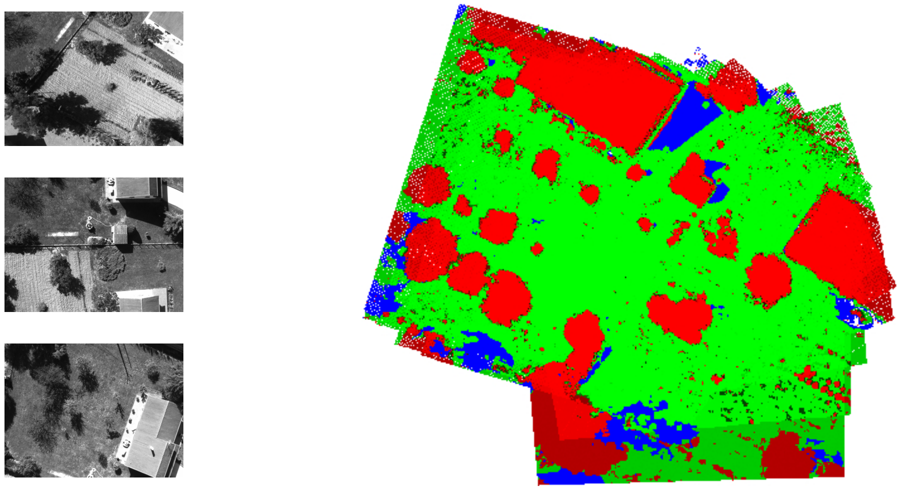

Fig. 1. Illustration of a traversability map built from a sequence of aerial monocular images [BLC06]. Left: some of the images processed, right: traversability map. Green, red and blue areas respectively correspond to traversable, obstacle and unknown ares.

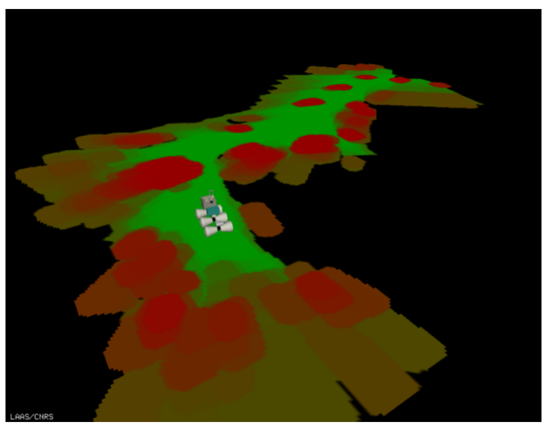

Fig. 2. A traversability map built by a ground rover as it navigates in the environment.

3 Perception for Air/Ground robotics systems

Integrating data perceived by aerial and ground robots is a multisensor fusion problem, that also encompasses the integration of the available initial data gathered by spatial or high altitude aerial means (such data being structured in a Geographic Information System – Fig. 3).

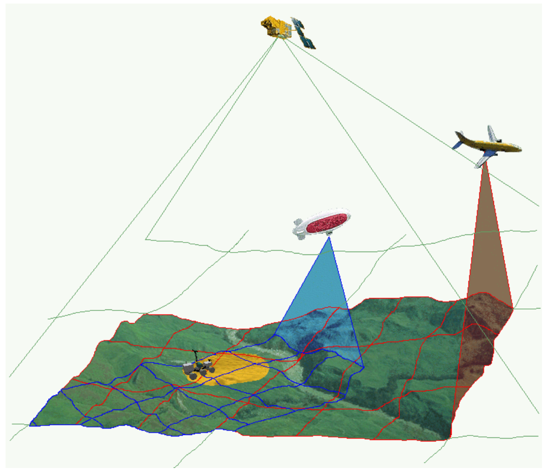

Fig. 3. A schematic view of the "multi-source" environment information integration problem among air and ground vehicles.

For that purpose, one must define algorithms and representations that support the following characteristics of the data:

- Data types: the acquired data can be images, either panchromatic or color, from which geometric 3D informations can be recovered, or directly 3D, as provided by a SAR or a Lidar for instance.
- Data resolution: the resolution of the gathered information can significantly change, depending on whether it has been acquired by a ground or an aerial sensor.
- Uncertainties: there can be several orders of magnitude of variation on data uncertainties between ground an aerial data.
- Viewpoints changes: beside the resolution and uncertainties properties of the sensors that can influence the detection of specific environment features, the difference of viewpoints between ground and aerial sensors generates occlusions that considerably change the effectively perceived area, and therefore the detectable features.

Fig. 3 illustrates some of these characteristics for a depth sensor (stereo or Lidar imagery). Of course, the difficulties to integrate the data depends on the considered contexts and sensors. For instance, for aerial sensors, occlusions caused by overheads are very unlikely to occur in open areas, whereas the presence of vegetation or buildings in urban areas is more challenging. Also, the UAV altitude strongly influences the properties of the perceived data in terms of precision and resolution, and can hinder the acquisition of range data. Note that among the possible sensors that can be embarked on a UAV, cameras have numerous advantages: images carry a vast amount of information on the environment regardless of the UAV altitude, they can also be used in micro-drones, and there is a tremendous amount of work available in the literature on environment modeling from aerial imagery.

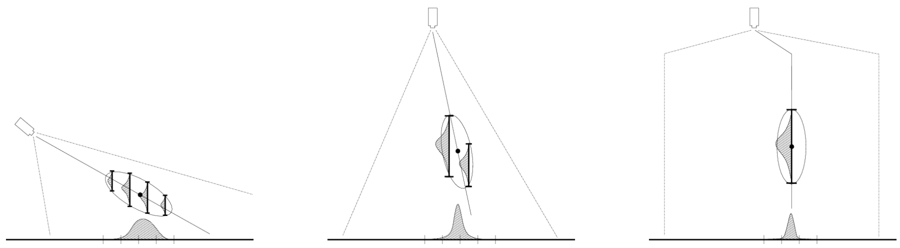

Fig. 4. 2D illustration of the viewpoint influence to derive a digital terrain map from depth imagery. Most of the uncertainty being in the depth estimate, encoding this uncertainty in a DTM requires different approaches with data perceived by a ground robot (left), a low altitude UAV (center) or a high altitude UAV (right). One can also see that the variable resolution of the data will play an important role in the specification of the processing algorithms.

3.1 Required models

The autonomous operation of the robots call for the structuring of the perceived data into a set of models, dedicated for specific purposes (Fig. 5):

- The evaluation of motions requires *traversability models*[4], that are confronted with the robot motion models in order to define goals and plan trajectories for the robots,
- In order to select viewpoints, *3D geometric models* are required to assess visibilities (between robots, between robots and targets, or between robots and specific features in the environment).
- Either for UAVs or UGVs, some motions can be defined with respect to environment features (*e.g.* servoing along a track or a straight line in the environment), detected by one or tho other kind of robot. Such features (denoted as *navigation supports*) also constitute an environment model.

And of course, two essential aspects lie behind the construction of such models:

- They must be *spatially consistent*, so as to faithfully represent the environment topology. This consistency is ensured thanks to Simultaneous Localization and Mapping (SLAM) approaches, that build and require *landmark maps*. Landmarks are sensor and robot dependant, but as pointed out in

[4] The word "traversability" is rather used for UGVs, but can be extended for UAVs, for which obstacles define no-fly volumes in the environment.

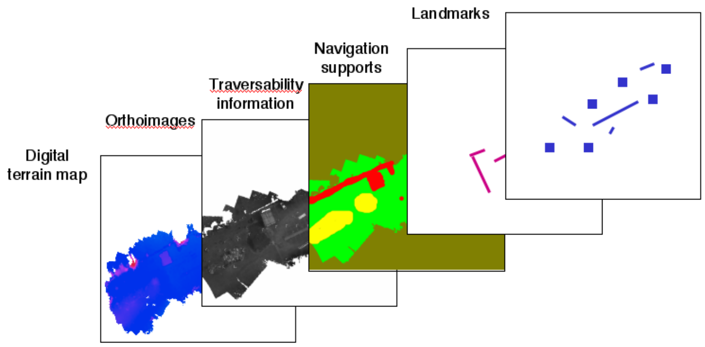

Fig. 5. The various environment models required are organized in a layered structure akin to a GIS. Some models are built and exploited by a single kind of robot, some are built by one kind of robot and used by the other one, whereas some result from the fusion of data gathered by both the UAVs and UGVs.

the example scenario of section 2.3, there is a need to share landmark representations to fuse the data acquired by the two kinds of robots.

- All the models must explicitly contain *the amount of information* they encode, be it the precision of geometric information or the confidence of semantic information: this is essential to plan data acquisitions, and therefore to define cooperating behaviours between the robots. This is the case even for missions that are not information oriented: for instance, in our simple Flying Eye scenario, it is the confidence on the estimated traversability that drives the UAV acquisitions – and therefore its motions.

3.2 Fusing aerial and ground data

If the literature provides numerous approaches to model the environment, be it from aerial of ground data, there are very few contributions that integrate both kind of data into consistent environment models. The fact is that the problem challenges the two essential functions required by environment modelling, namely *data registration* and *data fusion*.

Fig. 6 shows three different scenes perceived by a UAV and a UGV: in the leftmost case, the terrain is open and exhibit no structure. In the center images, some line segments can be quite easily matched between both images. In the rightmost images, the boardwalk is the only geometric feature that is perceived by both robots, whereas the detection of the white wrecked car should help to associate data (provided the white overexposed region is correctly identified as a car in the aerial data !). These examples show that there are many possible strategies achieve the data registration and fusion functions, that depend greatly on the environment context – and on the sensors used. We discuss some of these strategies below.

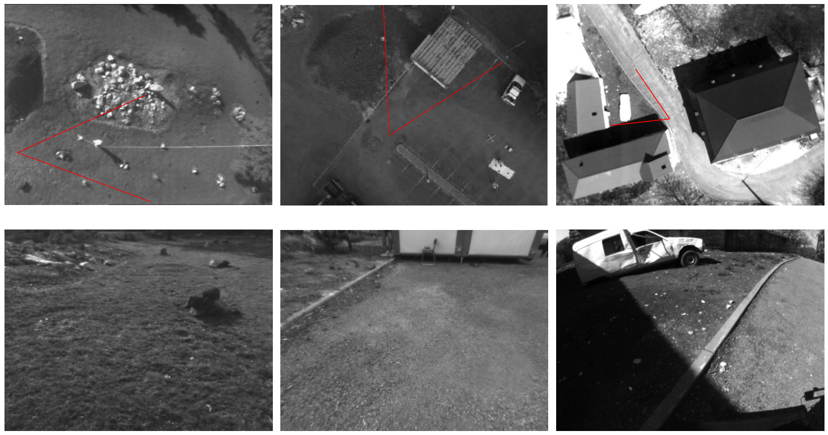

Fig. 6. Three different scenes perceived by an aerial robot (top), and a ground robot (bottom). The field of view of the ground robot is shown by red lines on the aerial images: even for a human operator, finding out the ground robot position and viewing direction in the aerial image is not an easy task.

Raw data registration

It can be readily seen from the images of Fig. 6 that not image feature based matching algorithm will help to register aerial and ground data. Nevertheless, 3D raw data can be registered using minimization processes (*e.g.* the iterative closest point algorithm, as in [DMH03, MHM05]), or by defining specific features that are be detectable in both kinds of data (*e.g.* the DTM registration approach presented in [VDH06] – DTMs are not strictly speaking raw data, but the process of building a DTM is rather a structuring process that resamples the data than a segmentation, classification or identification process). These approaches rely on the availability of Lidars, which can not be taken for granted in all application cases, and especially not with micro-drones.

Registering DTMs built from vision is a bit more challenging, mainly because the large imprecision of the depth data (that grows quadratically with the depth) yields less precise DTMs. Although in such cases the DTMs can be enriched with luminance information, one can see in Fig. 7 that hardly no stable features can be associated between the two maps.

SLAM with aerial and ground data

When considered in a SLAM context, the two registration and fusion functions are instantiated by the data association and the estimation functions.

Solutions to the SLAM problem have now reached a mature state, and have been successfully applied in various field robotics contexts [JFR07], for both aerial and ground robots. In particular, numerous efficient vision-based SLAM approaches have been recently proposed. But to our knowledge, no

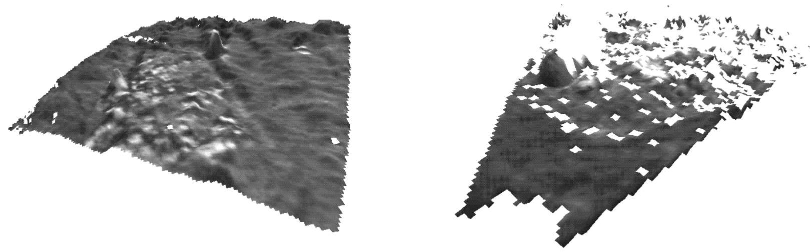

Fig. 7. Two local DTMs of the rocky area that can be seen in the leftmost images of Fig. 6: built from a single stereo image pair acquired at an altitude of about 20 meters (left), and from a stereo bench on board a rover (right). Both DTMs have the same resolution.

contribution tackle the air/ground data context: the data association problem is here very challenging.

As previously noted, there are little chances that image features often used in SLAM approaches (*i.e.* Harris or SIFT points) could be used to associate aerial and ground data. We rather believe that the solution lie in the use of higher level geometric primitives [LLB07]: 3D facets [MDR04], and more likely lines segments [ED06, LL07] or even planar patches [SMR07], as such features can be detected from both aerial and ground robot. Yet, these recent contributions to vision-based SLAM approaches call for robust data association schemes to be efficient for air/ground data fusion. For that purpose, geometric features could be enriched with additional information extracted in the images, as proposed for the object recognition problem in the vision community [RLSP03, LPS05].

Exploiting initial information

Exploiting initial information available on the environment is appealing. One of the main advantage is that such information are spatially consistent, and thus can avoid the use of SLAM approaches. But as in SLAM, the problem to solve is data association, and as in SLAM, we advocate that the use of high level geometric primitives, enriched by other qualitative attributes is the solution to look for.

Note that in the area of 3D GIS augmentation with ground vision data, there are a number of contributions worth to refer to, such as [FZ03] and [GA04]. But these approaches require that a fine geometric 3D model is built beforehand. On the contrary, [YM07] only require a 2D map that roughly describes the shape and position of buildings to localize a ground robot, using various features extracted from stereovision data.

4 Conclusion

Convincing demonstrations of aerial and ground robots that cooperate to fulfill a mission are yet to be achieved: the key problem to solve lie in the ability to properly register and fuse data obtained by both kinds of robots. It is a challenging issue, as one must define algorithms that are able to cope with the very different characteristics of the data, not to mention that they have to run under stringent real-time constraints.

We advocate that the use of high level geometric primitives is a prerequisite to tackle this issue, as shown by recent promising contributions to the SLAM problem. The ability to rely on semantic information would definitely be of a good help too, as labeled objects yield straightforward data associations. Even if such associations would not provide the necessary precision to register environment models at the precision of the required models, they can focus geometric features associations.

References

[BLC06] S. Bosch, S. Lacroix, and F. Caballero. Autonomous detection of safe landing areas for an UAV from monocular images. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing (China)*, Oct. 2006.

[CCG+05] L. Chaimowicz, A. Cowley, D. Gomez-Ibanez B. Grocholsky, M. Hsieh, H. Hsu, J. Keller, V. Kumar, R. Swaminathan, and C. Taylor. Deploying air-ground multi-robot teams in urban environments. In *International Workshop on Multi-Robot Systems, Washington, DC (USA)*, 2005.

[CGK+04] L. Chaimowicz, B. Grocholsky, J. Keller, V. Kumar, and C. Taylor. Experiments in multirobot air-ground coordination. In *IEEE International Conference on Robotics and Automation, New Orleans, LA (USA)*, 2004.

[DMH03] A. Downs, R. Madhavan, and T. Hong. Registration of range data from unmanned aerial and ground vehicles. In *32nd Applied Imagery Pattern Recognition Workshop*, pages 45– 50, Oct. 2003.

[ED06] E. Eade and T. Drummond. Edge landmarks in monocular slam. In *British Machine Vision Conference, Edinburgh (UK)*, Sep. 2006.

[FZ03] C. Fruh and A. Zakhor. Constructing 3D city models by merging aerial and ground views. *Computer Graphics and Applications*, 23(6):52–61, Nov-Dec. 2003.

[GA04] A. Georgiev and P.K. Allen. Localization methods for a mobile robot in urban environments. *IEEE Transactions on Robotics and Automation*, 20(5):851–864, Oct. 2004.

[GBK+04] B. Grocholsky, S. Bayraktar, V. Kumar, C.J. Taylor, and G. Pappas. Synergies in feature localization by air-ground robot teams. In *IEEE International Conference on Robotics and Automation, New Orleans, LA (USA)*, 2004.

[HBB+05] J. Harbour, S. Bauer, D. Bruemmer, D. Carroll, E. Pacis, K.D. Mullens, and H.R. Everett. Enabling technologies for unmanned protection systems. In *SPIE Proc. 5804 : Unmanned Ground Vehicle Technology VII, Orlando, FL (USA)*, 2005.

[JFR07] Special issue on SLAM in the field. *Journal of Field Robotics*, 11(1-2), Jan-Feb 2007.

[Joh00] A. Johnson. Surface landmark selection and matching in natural terrain. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

[LL07] T. Lemaire and S. Lacroix. Monocular-vision based SLAM using line segments. In *IEEE International Conference on Robotics and Automation, Roma (Italy)*, April 2007.

[LLB07] S. Lacroix, T. Lemaire, and C. Berger. More vision for SLAM. In *IEEE International Conference on Robotics and Automation, Roma (Italy), workshop on Unifying Perspectives in Computational and Robot Vision*, April 2007.

[LPS05] S. Lazebnik, J. Ponce, and C. Schmid. A sparce texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1265–1278, Aug 2005.

[MDR04] N. Molton, A.J. Davison, and I. Reid. Locally planar patch features for real-time structure from motion. In *British Machine Vision Conference, London (UK)*, Sept. 2004.

[MHM05] R. Madhavan, T. Hong, and E. Messina. Temporal range registration for unmanned ground and aerial vehicles. *Journal of Intelligent Robotics Systems*, 44(1):47–69, 2005.

[MTW+06] K. Mullens, B. Troyer, R. Wade, B. Skibba, and M. Dunn. Collaborative engagement experiment. In *SPIE Proc. 6230 : Unmanned Systems Technology VIII, Defense Security Symposium, Orlando, FL (USA)*, 2006.

[pro05] The CROMAT project. http://grvc.us.es/cromat/, 2003-2005.

[RLSP03] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. 3d object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. In *IEEE Conference on Computer Vision and Pattern Recognition, Madison, WI (USA)*, pages 272–277, June 2003.

[RNSDW02] M. Ridley, E. Nettleton, S. Sukkarieh, and H. Durrant-Whyte. Tracking in decentralised air-ground sensing networks. In *Fifth International Conference on Information Fusion*, July 2002.

[SKHR02] A. Stentz, A. Kelly, H. Herman, and P. Rander. Integrated air/ground vehicle system for semi-autonomous off-road navigation. In *AUVSIs Symposium on Unmanned Systems*, 2002.

[SMR07] G. Silveira, E. Malis, and P. Rives. An efficient direct method for improving visual SLAM. In *IEEE International Conference on Robotics and Automation, Roma (Italy)*, April 2007.

[VDH06] N. Vandapel, R.R. Donamukkala, and M. Hebert. Unmanned ground vehicle navigation using aerial ladar data. *International Journal of Robotics Research*, 25(1):31–51, Jan. 2006.

[YM07] J. Yun and J. Miura. Multi-hypothesis outdoor localization using multiple visual features with a rough map. In *IEEE International Conference on Robotics and Automation, Roma (Italy)*, 2007.