Vision-aided Inertial Navigation for Power Line Inspection

Jean-Philippe Tardif Michael George Michel Laverne
Alonzo Kelly Anthony Stentz
{*tardifj,mgeorge,mlaverne,alonzo,axs*}@*rec.ri.cmu.edu*
National Robotics Engineering Center, Carnegie Mellon University,
Ten 40th Street Pittsburgh, PA 15201, USA.

***Abstract*—We present a stereo vision-aided inertial navigation system and demonstrate its potential in power line inspection at close range using an unmaned aerial vehicle. This is made possible by recent developments in visual odometry and a newly proposed algorithm for the loose coupling of an inertial measurement unit and visual odometry. Our experiments show promising results.**

***Index Terms*—Visual Odometry, Inertial Navigation, Inspection, Unmanned Aerial Vehicle**

I. INTRODUCTION

THE inspection of electric transmission networks is of great importance for the power industry. It is an expensive and fastidious task not only because power lines cover a vast territory, but also because a significant portion of them are located in remote and hazardous environments. Furthermore, for safety reasons, human intervention should be reduced as much as possible. Almost 20 years ago, it was proposed to perform inspection using autonomous or remotely controlled Unmanned Aerial Vehicles (UAVs) [11], [15], [30]. However, technological limitations have slowed down their adoption. Today, commercial UAVs are remotely controlled and require their operators to be on site with the UAV when performing operations at close range like power line inspection. To overcome that limitation, UAVs should be completely autonomous or provide the remote operator with accurate information about its location and surroundings. In both cases, an accurate localization system is required.

The localization problem can be split into two aspects. Global pose is concerned with the estimation of the location and orientation of the vehicle with regard to a global coordinate system such as the one used by the Global Positioning System (GPS). It involves such sensors as Global Navigation Satellite System (GNSS) receivers and magnetic compasses. The current state of the art in localization for autonomous navigation relies on a combination of expensive sensors, notably Real Time Kinematic Global Positioning System (RTK-GPS) to achieve centimeter accuracy. Compared to conventional GPS, this, however, requires additional ground infrastructure that is expensive to deploy to cover very large areas.

Local pose, on the other hand, is concerned with the position and orientation of the vehicle with respect to the environment immediately surrounding it. In the robotics community, Simultaneous Localization and Mapping (SLAM) is a typical solution to this problem. In this work, we present a vision-aided inertial navigation system that is useful for local pose estimation. It is based on a stereo camera pair and an Inertial Measurement Unit (IMU) that can be mounted on any kind of wheeled, legged or aerial vehicle. Work is under way to incorporate a low cost GPS receiver in our system to provide accurate global pose as well, but it is not the topic of this work.

In the next section, we present some relevant prior work, followed by our vision-aided inertial navigation system in Section III. Proof of concept experiments are presented in Section IV followed by the conclusion in Section V.

II. PREVIOUS WORK

In this section, we review prior art on navigation systems based on vision and inertial sensors with potential application to UAVs.

At medium to high altitude, a common approach for motion estimation of a UAV is to successively align ground plane images [2], [20] using homographies. Complementary to this approach, Demonceaux *et al.* use an omnidirectional camera to detect the horizon line and estimate the altitude of the UAV [4]. Also, Rathinam *et al.* propose to follow locally linear structures such as power lines using visual feedback [23]. Those approaches are mostly suitable for long and smooth trajectories with objects of interest (typically the ground) located far from the UAV. For inspection at close range, a more general vision-aided inertial navigation system offers the advantage of estimating the motion of the UAV with respect to the inspected object. In addition, a permanent view of the ground is not necessary allowing more flexibility in the motion of the UAV.

There have been several aided-inertial navigation systems introduced in the literature with different formulations. These are generally influenced by the choice of SLAM algorithm and the motion model. For example, when employing Extended Kalman Filter (EKF) based SLAM, it is natural to combine it with the inertial EKF to jointly estimate the sensor states and visual landmarks [17], [21], [22], [26]. Mourikis *et al.* demonstrated such an approach for planetary landing applications [18]. Other approaches rely on a modified EKF to

This material is based upon research supported by USDA/CSREES under Award Number 2008-51180-04890.

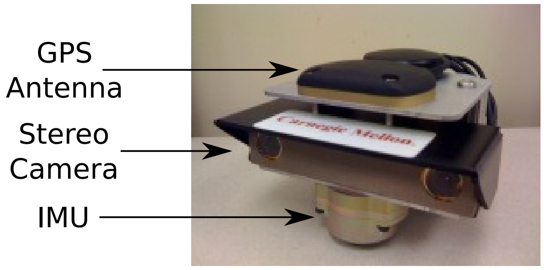


Fig. 1. Our positioning sensors. We treat the stereo camera and IMU as collocated with a fixed and known rotation. The GPS is only used for ground truth comparison.

use constraints computed on pairs of consecutive images [5], [12], [24]. Doing so guarantees that vision measurements are not correlated, at the expense of less accurate visual odometry (VO).

In the computer vision community, structure-from-motion and visual odometry are the counterparts of SLAM in the robotics community. Rather than using a filtering formulation, VO formulates the localization task as an optimization problem and impressive results have been demonstrated in recent years [13], [14], [19], [28]. In their work, Konolige *et al.* argue that ignoring correlation is a price worth paying for relying on VO rather than SLAM [14]. They propose a loosely coupled system which combines two EKFs and stereo VO. They rely on a cascaded EKF where a low-level EKF is used to process inertial measurements and a high-level EKF fuses VO with filtered inertial measurements. This allows them to perform predictions using the VO and corrections with the IMU. We adopt the opposite approach and predict with the IMU measurements in our EKF, since the IMU has the highest update frequency. Furthermore, their approach assumes that acceleration profile is of zero mean locally in order to estimate roll and pitch angles. As a consequence, the uncertainty needs to be artificially high to compensate for this assumption. In our formulation, the direction of gravity implicitly damps roll and pitch errors without any arbitrary assumptions. As a result, the computed uncertainties are closer to the true uncertainties.

III. Vision-aided Inertial Navigation System

Building a motion estimation system by combining an exteroceptive and a proprioceptive sensor is an attractive solution because, roughly speaking, they have opposed strengths and weaknesses. The former estimates motion based on external observations, for example, using video cameras. As a result, error accumulation or drift is essentially proportional to the length of the trajectory. The drawback is that accuracy is also dependent on the geometry of the environment. Proprioceptive sensors, by their nature, measure their own motion. Thus, they can operate in any kind of environment. However, error accumulation is a function of time rather than distance which is why they require some form of aiding.

Our navigation system consists of two loosely coupled components. The main component is a modification to the conventional inertial navigation Extended Kalman Filter and the second is a stereo based VO. The filter is used to fuse the inertial data and the motion estimates provided by the VO. In the following, we first describe our VO algorithm followed by our modified EKF. A more in depth description is available in [27].

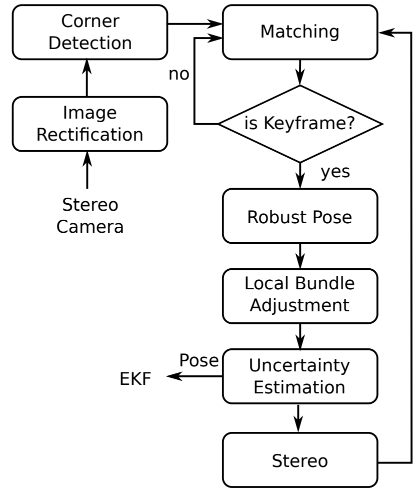


Fig. 2. Flow chart summarizing our visual odometry system, see Section III-A for details. A flow chart of the EKF is shown in Figure 3.

A. Visual Odometry

Historically, stereo cameras have been used as range sensors and pose estimation was performed by successive alignment of the current and previous point clouds with the Iterative Closest Point (ICP) algorithm [10], [16]. Similarly to Nister *et al.* and Konolige *et al.* , our VO instead relies on incremental Structure From Motion (SFM) [1], [19]. In this context, there is no fundamental difference between using a single camera or a stereo camera. The additional camera of the stereo pair simply provides additional information during motion estimation and measurements for landmark triangulation. In SFM, contrary to ICP-based solutions, the position of the observed landmarks are estimated using all frames in which they appear. Besides increased accuracy, this makes pose estimation possible even when the landmarks are located far from the cameras and have virtually zero image disparity. In addition, relying on a stereo camera considerably reduces drift in the scale (or speed) of the computed motion. Finally, initialization is simplified because the depth of the landmarks can be instantly estimated. A description of our algorithm follows and a flow chart is given in Figure 2.

Our stereo camera is calibrated offline, a critical step for an accurate and unbiased trajectory estimation. Images are simultaneously obtained from both cameras and rectified to horizontally align their epipolar lines. Features are detected with sub-pixel accuracy in both images using the Harris corner detector. After the first frame is obtained, sparse stereo matching is performed: features from the left and right images are matched using normalized cross correlation. Matched features can then be triangulated yielding the first set of landmarks. Thus, some disparity must appear in the image for proper initialization of the system.

Every time a new set of stereo images is acquired, motion

Steps	Computation time (milliseconds)
Image rectification	9
Corner extraction	5
Feature matching	10
Sparse stereo matching∗	1
Robust pose estimation*	5
Bundle adjustment*	20

TABLE I
COMPUTATION TIME ON AN INTEL CORE 2 DUO PROCESSOR. STEPS MARKED WITH A ∗ ARE PERFORMED ONLY AT KEY FRAMES.

estimation is performed as follows. Features corresponding to 3D landmarks are matched with features from the new stereo pair. Note that we rely on *matching* with mutual consistency rather than *tracking* on correlation windows with some rejection threshold or the Lucas Kanade tracker. This results in a set of 2D-3D correspondences used to robustly estimate the current position and orientation of the cameras. By *robust*, we mean that it can perform pose estimation despite a large porportion of wrong correspondences due to matching errors. We use the 3-point algorithm [9] in conjunction with Random Sampling and Consensus (RANSAC) [8] where the error criterion is the reprojection error of the landmarks in both stereo images. This is followed by iterative refinement. In addition, we perform local bundle adjustment which simultaneously refines the last few pose estimates of the stereo pair and the currently observed landmarks [29]. This is known to significantly increase the accuracy of the estimated trajectory [6], [14], [25]. Erroneous landmarks are rejected and replaced by new ones using sparse stereo.

Contrary to conventional SLAM approaches [3], pose estimation and landmark triangulation is only performed on a subset of frames called *key frames*.This allows real time computation at 30 frames per second despite expensive nonlinear optimization techniques. Typical computation times for each step are given in Table I. Observe that for a stereo camera running at 30 Hz, pose estimation at every frame is impossible. However, real-time computation is feasible if a key-frame appears no more than every three frames. In addition, we do not rely on a motion model or any other kind of smoothing. Six degrees of freedom motion estimates are obtained at each key-frame.

VO provides the relative pose between the current key frame and the previous key frame to the EKF along with its uncertainty. Unlike the uncertainty of the current pose in the global frame [7] this does not require propagating uncertainties in time. The uncertainty of the relative pose is a function of the currently observed landmarks only. In case of failure, the uncertainty is simply set to infinity and the current pose set to identity (anything can in fact be used). For the EKF, this is easy to handle as explained in the next section. Failure of the VO only happens when the number of image features is almost null as a result of high motion blur, over or under image exposure or if the observed environments simply don't have any salient regions.

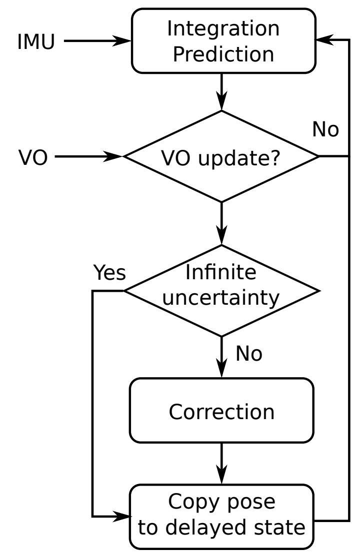


Fig. 3. Flow chart of the EKF taking inputs from the IMU as well as from the VO (Figure 2). Details in Section III-B.

B. Fusion

The Extended Kalman filter estimates the current system state: position, velocity and orientation in three dimensions as well as some calibration and noise shaping states. This structure allows an Inertial Measurement Unit (IMU) to predict the state at high frequency and be corrected and calibrated online at lower frequencies by the VO. A flow chart summarizing the steps taken by EKF is given in Figure 3.

To incorporate VO relative pose updates, the Kalman filter uses a delayed state formulation where the filter states are:

- UAV's current position and orientation (6 dof);
- gyroscope biases (3 dof);
- accelerometer biases (3 dof)
- UAV's current velocity (3 dof);
- UAV's position and orientation at the previous VO update (6 dof, the 'delayed states').

Similar formulations have been referred to in the literature as stochastic cloning [17]. This structure allows natural incorporation of the VO measurement in a loosely coupled arrangement.

At initialization, the filter's position, orientation and the delayed states are identical. A global vertical reference is provided by gravity measurement but the system does not necessarily have any knowledge of its absolute global heading.

With each IMU measurement, integration is performed so the states of the filter are updated with the exception of the delayed states that remains unchanged, that is, no smoothing is performed. Correction is performed when the VO outputs a relative pose between the current position/orientation of the UAV and the previous delayed states. The measurement corrects errors introduced through the IMU particularly. After a step of correction, the current state is copied onto the delayed states of the vehicle. Thus, they are briefly identical, before the IMU again begins to track the subsequent system motion. Care is taken to handle the associated uncertainties correctly.

When the VO fails to estimate the pose of the UAV, the

Fig. 4. We simulated a UAV flight using a hand-held mono-pod (red circle) and 15 meter platform lift.

EKF will receive a relative pose of infinite uncertainty. In this case, no correction is performed and the current states of the EKF are copied to the delayed states as the next VO update will be given with respect to that pose.

IV. Proof of concept experiments

Our system, shown in Figure 1, consists of a Honeywell HG1930 IMU mounted underneath a PointGrey Bumblebee 2 stereo camera running at 30 Hz with a 12 centimeter baseline and a field of view of 65 degrees. In addition, we use a Novatel OEMV-3 GPS receiver to acquire ground truth position for comparison. Note that conventional GPS is accurate only to a few meters which is not enough for our application. So we obtain centimeter accuracy by post-processing the GPS data using Novatel proprietary software and publicly available CORS and IGS reference stations (the data from these reference stations are made public on the day following their collection). Our computer is a conventional PC equipped with an Intel Core 2 Duo clocked at 2.4 Ghz. In our system, visual odometry is by far the most computationally expensive component. The number of tracked features is automatically adjusted to keep the computation in real-time whilst leaving enough overhead to run the EKF.

We simulated a remotely controlled UAV by attaching our navigation system to the tip of a hand held 1.5 meters mono-pod. We placed a platform lift that could go as high as 15 meters next to a power line pole. Using the lift and hand-held positioning sensors, we simulated the taking off and ascent of a UAV next to the pole, inspection of the power line and transformer followed by the descent of the UAV. Although our system can process the data in real-time, data from the cameras, IMU and GPS receiver were only recorded and processed off-line. Pictures of the acquisition process are shown in Figure 4. We acquired three datasets with trajectories of length varying between 15 and 22 meters. This resulted in between 1654 and 2532 key frames per dataset. The longest of our datasets is shown in Figure 5. In this figure, only the landmarks close to the INS are shown. In fact, a lot more are recovered from the surrounding environments, particularly from trees and buildings.

Naturally, the estimated trajectories suffer from drift in both position and orientation. To assess the accuracy of the estimated trajectories, we compare them to the post-processed GPS. Trajectories given by INS and GPS are shown in Figure 6. The separation between the trajectories are also given in

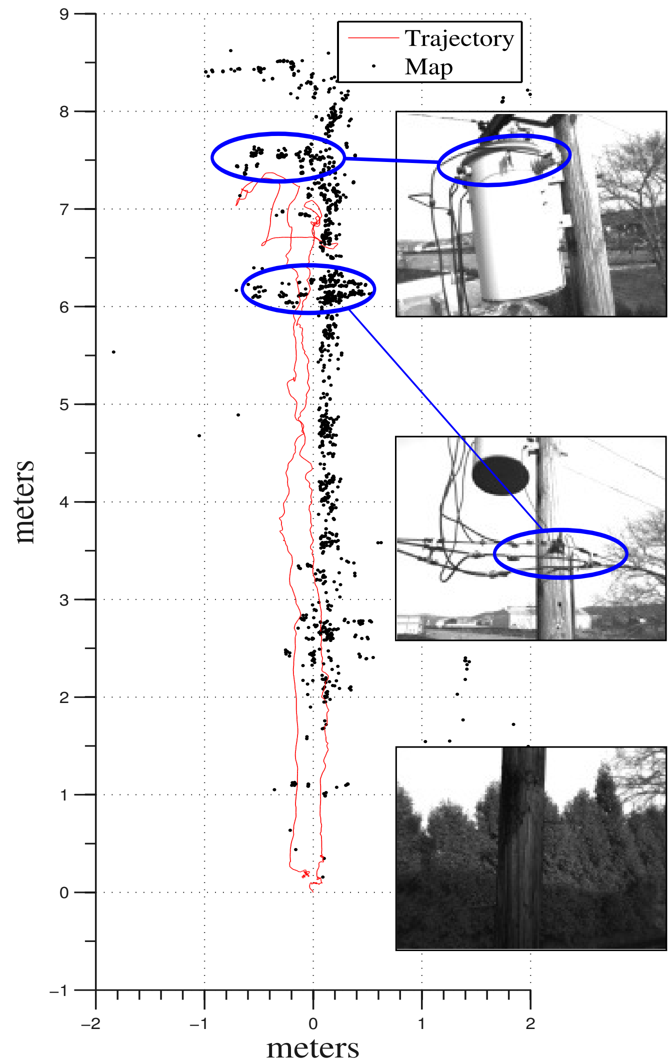
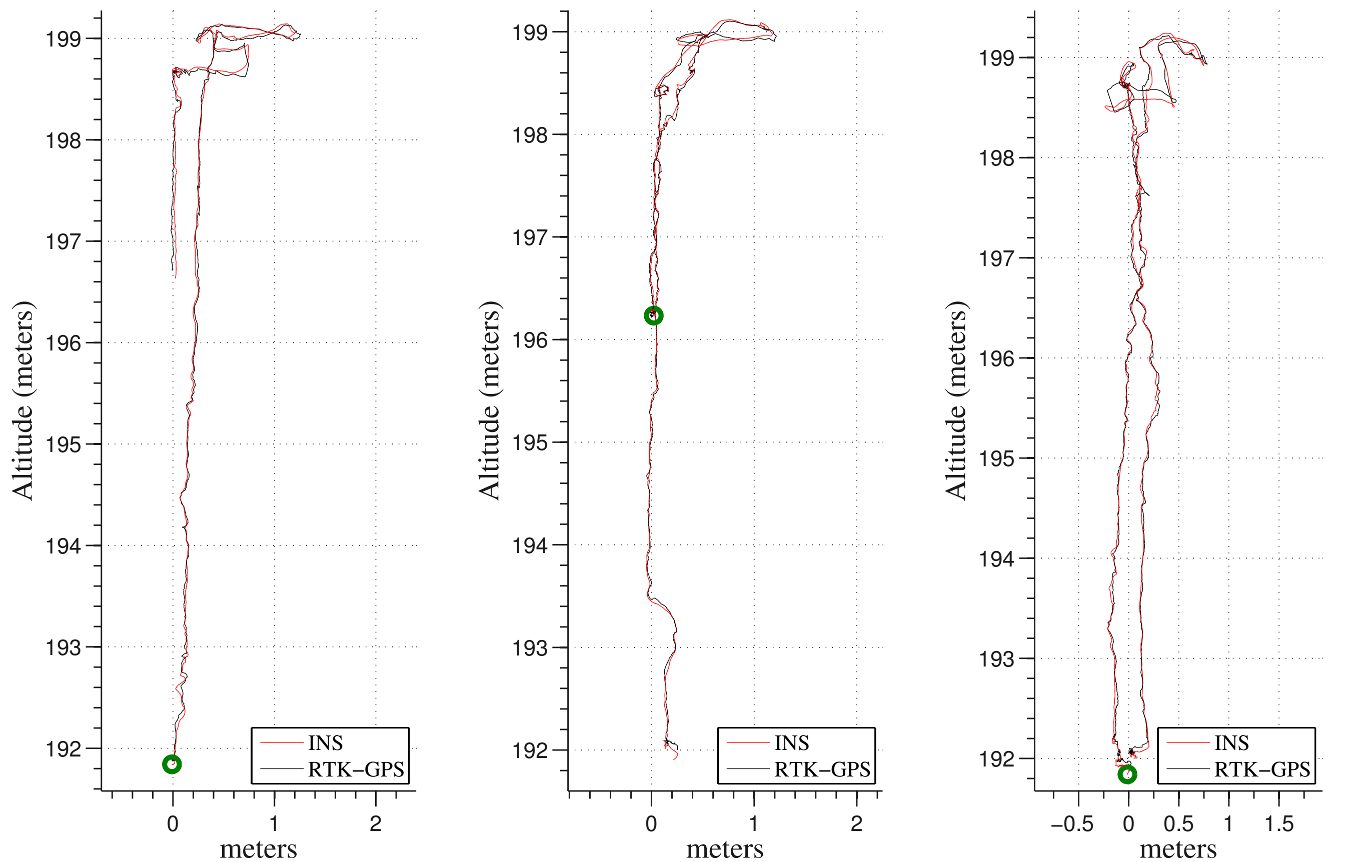


Fig. 5. Recovered trajectory and point cloud map from our third datasets. Only a subset of the point cloud close to the INS (below 3 meters) is shown. Sample images are from the left camera.

Figure 7. In the case of our longest dataset, the position error is at most 20 cm. Note that the initial orientation of the INS is only accurate to a few degrees. Such inaccuracies artificially increase the perceived amount of drift of the estimated trajectories. Since our goal is to quantify the accuracy of our local pose estimate, we correct our initial orientation by minimizing the alignment between the GPS and estimated trajectory in the first 2 meters traveled. In general, this decreases the maximum error by around 10 centimeters.

V. Conclusion and Future work

In this paper, we demonstrated the potential of a vision-aided inertial navigation system that could be mounted on a UAV. It provides an accurate local pose as well as point cloud map of the environments. In its current form the system would most useful for visualization of a remotely controlled UAV. Current work focuses on combining our INS with various types of autonomous vehicles performing path following. In addition, we are working towards integrating a conventional



(a) Dataset 1 (b) Dataset 2 (c) Dataset 3

Fig. 6. Comparison between the three estimated trajectories and GPS. The length of the trajectories varies between 15 and 22 meters. The starting point is marked by a green circle. The maximal error is around 20 centimeters (see Figure 7 for details).

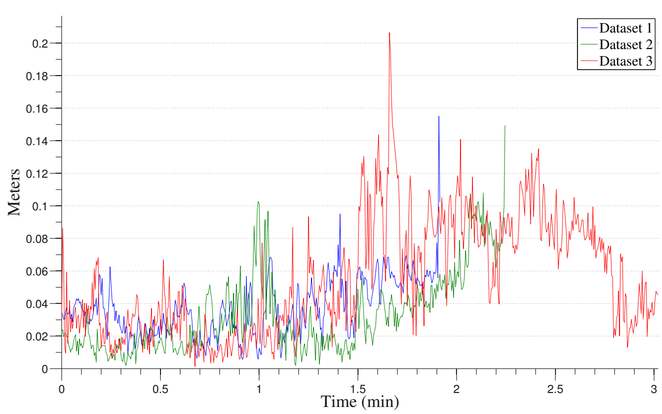


Fig. 7. Error of the estimated trajectory with respect to GPS for our three datasets.

low cost GPS to reduce long term drift and provide accurate global pose.

References

[1] M. Agrawal and K. Konolige. Rough terrain visual odometry. In *International Conference on Advanced Robotics*, 2007.

[2] F. Caballero, L. Merino, J. Ferruz, and A. Ollero. Vision-based odometry and slam for medium and high altitude flying UAVs. *Unmanned Aircraft Systems*, page 137–161, 2008.

[3] A. J. Davison, I. D. Reid, N. Molton, and O. Stasse. MonoSLAM: Real-Time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, 2007.

[4] C. Demonceaux, P. Vasseur, and C. Pégard. Omnidirectional vision on UAV for attitude computation. In *Proceedings 2006 IEEE International Conference on Robotics and Automation*, page 2842–2847, 2006.

[5] David D. Diel, Paul DeBitetto, and Seth Teller. Epipolar constraints for Vision-Aided inertial navigation. In *IEEE Workshop on Motion and Video Computing*, volume 2, pages 221–228, Los Alamitos, CA, USA, 2005. IEEE Computer Society.

[6] C. Engels, H. Stewenius, and D. Nister. Bundle adjustment rules. *Photogrammetric Computer Vision*, 2, 2006.

[7] A. Eudes and M. Lhuillier. Error propagations for local bundle adjustment. In *IEEE International Conference on Computer Vision And Pattern Recognition*, 2008.

[8] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[9] Haralick, Lee, Ottenberg, and Nolle. Review and analysis of solutions of the three point perspective pose estimation problem. *International Journal of Computer Vision*, 13(3):331–356, December 1994.

[10] A. Howard. Real-time stereo visual odometry for autonomous ground vehicles. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, page 3946–3952, Nice, France, 2008.

[11] D. Jones. Power line inspection - a UAV concept. *IEE Seminar Digests*, 2005(11271):6, 2005.

[12] Jonathan Kelly, Srikanth Saripalli, and Gaurav Sukhatme. Combined visual and inertial navigation for an unmanned aerial vehicle. In *Field and Service Robotics*, pages 255–264. 2008.

[13] K. Konolige and M. Agrawal. Frameslam: From bundle adjustment to real-time visual mapping. *IEEE Transactions on Robotics*, 24(5):1066–1077, 2008.

[14] K. Konolige, M. Agrawal, and J. Sola. Large scale visual odometry for rough terrain. In *International Symposium on Robotics Research*, 2007.
[15] L. Ma and Y. Q Chen. Aerial surveillance system for overhead power line inspection. *Center for Self-Organizing and Intelligent Systems (CSOIS), Utah State University, Tech. Rep. USU-CSOIS-TR-04-08*, 2003.
[16] L. H Matthies. *Dynamic stereo vision, Ph.D dissertation*. Carnegie Mellon University, 1989.
[17] A. I Mourikis and S. I Roumeliotis. On the treatment of relative-pose measurements for mobile robot localization. In *IEEE International Conference on Robotics and Automation*, page 15–19, 2006.
[18] A. I Mourikis, N. Trawny, S. I Roumeliotis, A. Johnson, and L. Matthies. Vision aided inertial navigation for precise planetary landing: Analysis and experiments. In *Proc. Robotics Systems and Science Conference*, 2007.
[19] D. Nister, O. Naroditsky, and J. Bergen. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23(1):3–20, 2006.
[20] K. Nordberg, P. Doherty, G. Farneback, P. E Forssén, G. Granlund, A. Moe, and J. Wiklund. Vision for a UAV helicopter. In *IROS, workshop on aerial robotics*, 2002.
[21] L. L. Ong, M. Ridley, J. H Kim, E. Nettleton, and S. Sukkarieh. Six DoF decentralised SLAM. In *Australasian Conference on Robotics and Automation*, page 10–16, 2003.
[22] G. Qian, R. Chellappa, and Q. Zheng. Robust structure from motion estimation using inertial data. *Journal of the Optical Society of America A*, 18(12):2982–2997, 2001.
[23] S. Rathinam and R. Sengupta. Vision-Based monitoring of locally linear structures using an unmanned aerial vehicle. *Journal of Infrastructure Systems*, 14:52, 2008.
[24] S. I Roumeliotis, A. E Johnson, and J. F Montgomery. Augmenting inertial navigation with image-based motion estimation. In *Proceedings-IEEE International Conference on Robotics and Automation*, volume 4, page 4326–4333, 2002.
[25] H. Strasdat, J. M. M. Montiel, and A. J Davison. Real-Time monocular SLAM: why filter? In *Int. Conf. on Robotics and Automation, Anckorage, AK*, 2010.
[26] D. Strelow and S. Singh. Motion estimation from image and inertial measurements. *The International Journal of Robotics Research*, 23(12):1157, 2004.
[27] J. Tardif, M. George, M. Laverne, A. Kelly, and A. Stentz. A new approach to vision-aided inertial navigation. In *International Conference on Intelligent Robots and Systems*, 2010.
[28] J. Tardif, Y. Pavlidis, and K. Daniilidis. Monocular visual odometry in urban environments using an omnidirectional camera. In *International Conference on Intelligent Robots and Systems*, 2008.
[29] B. Triggs, P. F McLauchlan, R. I Hartley, and A. W Fitzgibbon. Bundle adjustment-a modern synthesis. *Lecture Notes in Computer Science*, page 298–372, 1999.
[30] Carnegie Mellon University. Autonomous helicopter project, 1991.

Jean-Philippe Tardif is currently a post-doctoral fellow with the National Robotics Engineering Center at Carnegie Mellon University, under supervision of Prof. Tony Stentz. Previously, he was a post-doctoral fellow at McGill University and at the University of Pennsylvania. He received his Ph.D. in computer vision from Université de Montréal in 2007. Over the years, he has also collaborated with researchers from INRIA Rhône-Alpes and CNRS Clermont-Ferrand in France. His areas of research are structure from motion, visual odometry, omnidirectional camera calibration and self-calibration, structured light reconstruction and place recognition.

Michael George is a Robotics Engineer with the National Robotics Engineering Center (NREC) at Carnegie Mellon University (CMU). He received an MS in Robotics from the Australian Center for Field Robotics at the University of Sydney in 2007. He works on inertial navigation and sensor fusion for personal navigation and agricultural robotics.

Michel Laverne is a Robotics Engineer at Carnegie Mellon University's NREC in Pittsburgh, PA. He received his Master's degree in Computer Sciences and Modeling from ISIMA in Clermont-Ferrand, France in 2006. Before joining CMU, he also worked at the NIST in Gaithersburg, MD and for CERN in Geneva, Switzerland. His research interests include embedded systems, sensors and software for robotic application.

Alonzo Kelly is currently an associate professor at the Robotics Institute of Carnegie Mellon University. He has formerly held technical staff positions at Cal Tech's Jet Propulsion Lab and at MD Robotics. He received his PhD in Robotics from CMU in 1995. His research interests include all aspects of mobile robots including pose estimation, control, mapping, perception motion planning, simulation and modeling and user interfaces.

Anthony Stentz is a Research Professor at the Robotics Institute, Carnegie Mellon University, and Associate Director of the Robotics Institute's National Robotics Engineering Center. He received his Ph.D. in computer science from Carnegie Mellon University in 1989, his M.S. in computer science from CMU in 1984, and his B.S. in physics from Xavier University of Ohio in 1982.

Dr. Stentz's research expertise includes unmanned ground vehicles, unmanned air vehicles, dynamic planning, multi-vehicle planning and coordination, perception for mobile vehicles, robot architecture, and artificial intelligence in the context of fieldworthy systems. He has over 150 journal articles, conference and workshop papers, books, technical reports, and patents to his credit. Dr. Stentz is a recipient of the Alan Newell Award for Research Excellence for his agricultural automation work and a NASA Board Award for developing software used on the Mars Exploration Rovers.