STEREO VISUAL SYSTEM FOR AUTONOMOUS AIR VEHICLE NAVIGATION

Luis Mejias [*] **Pascual Campoy** [*] **Ivan Mondragón** [*]
Patrick Doherty [**]

[*] *Computer Vision Group*
Universidad Politécnica de Madrid
Madrid 28006 - Spain
`first.last@upm.es`
[**] *Department of Computer and Information Science*
Linköping University
Linköpings, Sweden
`patdo@ida.liu.se`

Abstract: We present a system to estimate the altitude and motion of an aerial vehicle using a stereo visual system. The system has been initially tested on a ground robot and the novelty lays on its application and robustness validation in an UAV, where vibrations and rapid environmental changes take place. The two main functionalities are height estimation and visual odometry. The system first detects and tracks salient points in the scene. Depth to the plane containing the features is calculated matching features between left and right images then using the disparity principle. Motion is recovered tracking pixels from one frame to the next one finding its visual displacement and resolving camera rotation and translation by a least-square method. We present results from different experimental trials on the two platforms comparing and discussing the results regarding the trajectories calculated by the visual odometry and the onboard helicopter state estimation.

Keywords: autonomous helicopter, vision-based navigation

1. INTRODUCTION

A key task in navigational applications is the robot self-localization. The use of computer vision for ground robot obstacle avoidance, odometry and navigation, in general, has been largely studied contributing with many solutions (Nistér *et al.*, 2006). Visual motion estimation for aerial robotics field, however, is becoming a important research area where researcher are investing significant effort.

UAV tasks in restricted environment at low altitudes where they are close to obstacles is difficult due to GPS dropouts. Flying at low altitudes in constrained environment makes the UAV more vulnerable and many problem may arise. Computer vision as passive sensor not only offer a rich source of information for navigational purposes, but it can be used as main sensor in such environment to substitute the GPS as navigational sensor. With the increasing interest in UAVs, a visual system that can be able to determine the robot 3D location in its operational environment is becoming a key sensor for civil applications. Motion estimation from imagery is often refereed as visual odometry (Matthies, 1989). its main idea is to determine vehicle's position and orientation by detecting and tracking visually salient points using an onboard camera. This technique has been traditionally used in ground vehicles to improve robot's localization. Its have been documented the use of visual odometry techniques on different mars rovers (Cheng *et al.*, 2006) and on a small Shrimp robot (Milella and Siegwart, 2006). In the case of aerial vehicles, one of the first stereo systems for an autonomous helicopter documented is described in (Amidi, 1996). This system used a customized vision processing hardware to estimate helicopter position. A comparison of two height estimation approaches using stereo and sonar is given in (Corke *et al.*, 2000).

The paper presents a stereo visual system to compute the altitude and relative displacement for an autonomous helicopter. The algorithm was tested in a ground robot to evaluate the functionalities before implement it onboard an autonomous helicopter. Results from experimental trials in both system are shown. Section 2 presents the visual algorithm for height and motion estimation. Section 3 describes the platform used as testbed and experimental setup. Experiments using the ground robot and the helicopter are described in section 4. Conclusions are presented in section 5.

2. VISUAL APPROACH

This section gives a detailed description of the visual techniques used in the system. The algorithm starts detecting salient points (corners) at each stereo frame. The corners detected are just salient features in the environment, no need for special landmark or specific features is required for the algorithm to work. In fact, it is able to detect and track features in a variety of terrain and environments. With the set of corners, first assuming a flat plane containing the features a disparity measure is obtained, then using a basic stereo disparity principle the distance to the plane that contains the features is computed. Secondly, using the features from one of the stereo images and tracking them from one frame to another the instantaneous visual displacement of the set of features is recovered. The visual feature displacements are used to find the camera motion using two additional operations as Iterative Closest Point Algorithm (ICP) combined with Single Value Decomposition (SVD) to find the camera transformation.

2.1 Feature Detection

We detect corners using a $Harris$ corner detector (Harris and Stephens, 1988). There exist numerous corners detector in the literature. Several analysis and comparatives have demonstrated the $Harris$ operator to be stable under small to moderated transformations (Ashbrook, 1992) being an appropriated choice for our purposes. This operator is based on the ideas of moravec (Moravec, 1977). These operators compute a local measure of cornerness C as a set of products of local gradient magnitudes and rate of change of gradient direction. Our actual implementation is based on the variant proposed by nobel (Noble, 1989) which define

$$C = \frac{\langle I_x^2 \rangle + \langle I_y^2 \rangle}{\langle I_x^2 \rangle \langle I_y^2 \rangle - \langle I_x I_y \rangle^2} \tag{1}$$

where I_x, I_y are the image derivatives in x and y using a convolution mask of the form g_x=[-1 -1 -1; 0 0 0; 1 1 1] and g_y=[-1 0 1; -1 0 1; -1 0 1]. I_x^2, I_y^2 are $I_x I_x$ and $I_y I_y$. $\langle I_x^2 \rangle = I_x^2 * G$, $\langle I_y^2 \rangle = I_y^2 * G$ and $\langle I_x I_y \rangle^2 = (I_x I_y * G)^2$ are the convolution with a gaussian smoothing kernel (G) of sigma σ over a window of size W=5 × 5 . A good corner is such that has a small value of C. The N lowest values of C are kept as salient features in the scene.

2.2 Stereo disparity for height estimation

The UAV height can be computed finding the distance to the plane containing the features, in our case, while the UAV is flying the features are detected over a flat ground thus the altitude is computed. In order to find this distance the *stereo disparity* principle is used, which is the difference in position between correspondence points in two images, left and right. Disparity is inversely proportional to scene depth multiplied by the focal lenght (f) and baseline (b). The depth is computed using the expression for Z shown in figure 1. To find the correspondence between features in left and right images we use zero mean normalized cross-correlation (ZNNC) which has advantages like good stability in the presence of changes in ambient light (Martin and Crowley, 1995). The process can be summarized as follows: Search at every detected feature for candidates of maximum correlation from left to right, and back right to left to check their consistency. Eliminate spurious candidates. Refine disparity to sub-pixel level. Estimate distance of points.

2.3 Pixel tracking for motion estimation

Feature tracking is performed using at first stage the same technique used for feature correspondence between left and right corners, the zero mean normalized cross-correlation (ZNNC). Correlation is performed within a certain pixel distance from each other keeping those points having a correlation coefficient higher than 0.85. Pixels

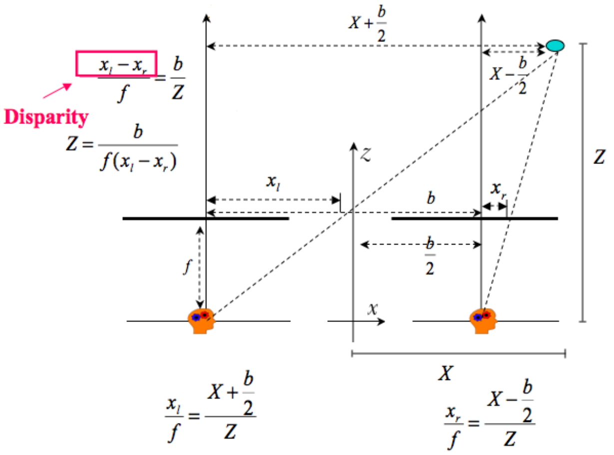


Fig. 1. Stereo disparity principle for parallel cameras.

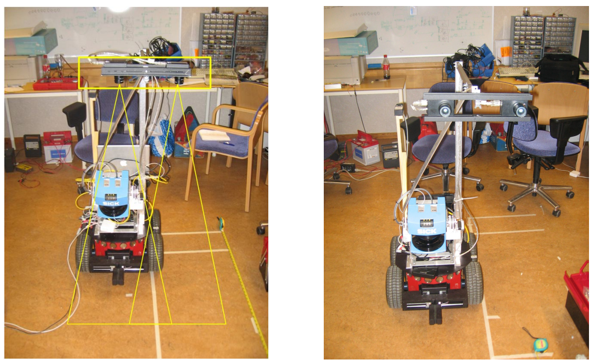
are tracked from one frame to the next. The motion problem estimation is done aligning two set of points which correspondence is known and finding the rotation matrix and translation vector, i.e, 3D transformation matrix T that minimizes the mean-squares objective function $min_{R,t} \sum_N \parallel TP_{k-1} - P_k \parallel^2$. We approach the problem using ICP registration and motion parameter estimation using SVD. Assume we have two set of points which we call data and model, $P = \{p_i\}_1^{N_p}$ and $M = \{m_i\}_1^{N_m}$ with $N_p \neq N_m$ and which correspondence is known. The problem is to compute the rotation (R) and translation (t) producing the best alignment of P and M, $M = RP + t$. Let us define the closest point in the model to a data point p as $cp(p) = arg\ min_{m \in M} \parallel m - p \parallel$, then we can summarize the ICP step as

(1) compute the subset of closest points (CP) , $y = \{m \in M \mid p \in P : m = cp(p)\}$
(2) compute the least-squares estimate of motion bringing P onto y, $(R, t) = argmin_{R,t} \sum_{i=1}^{N_p} \parallel y_i - Rp_i - t \parallel^2$
(3) apply motion to the data points, $P \leftarrow RP + t$
(4) if the stopping criterion is satisfied, exit; else goto 1.

Calculating the rotation and translation matrix using SVD can be summarize as follow: first, rotation matrix is calculated using the centroid of the set of points, centroids are calculated as $y_{c_i} = y_i - \bar{y}$ and $p_{c_i} = p_i - \bar{p}$, where $\bar{y} = \frac{1}{N_p} \sum_{N_p} cp(p_i)$ and $\bar{p} = \frac{1}{N_p} \sum_{N_p} p_i$, then rotation is found minimizing $min_R \sum_{N_p} \parallel y_{c_i} - Rp_{c_i} \parallel^2$, this eq. is minimized when trace(RK) is maximized with $K = \sum_{N_p} y_{c_i} p_{c_i}^T$. Matrix K is calculated using SVD as $K = VDU^T$, thus the optimal rotation matrix that maximizes the trace is $R = VU^T$. The optimal translation that aligns the centroids is $t = \bar{y} - P\bar{p}$. An example of the pixel tracking algorithm can be seen in (Mejías *et al.*, 2007).

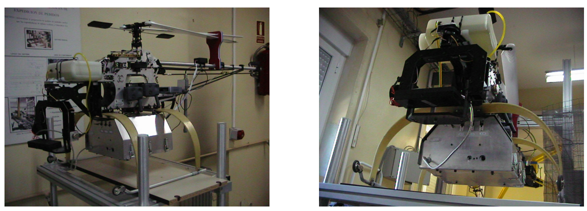
3. THE TESTBED AND EXPERIMENTAL TASK

We have tested the system on different platforms and environments. First, motion and height estimation functionalities were tried on a Pionner AT2 (figure 2) ground robot equipped and configured for indoor-outdoor experiments at University of Linköpings, Sweden. The robot was equipped with a videre STH stereo head attached to a PowerBook running Linux PPC for image processing and a GPS in single mode for outdoor experiments. Two camera configurations were used. Camera looking downward (figure 2a) to detect and track features over the ground to test odometry. Camera looking forward (figure 2b) to test depth estimation. Second, the system was configured and implemented on the COLIBRI (Mejías *et al.*, 2007) autonomous helicopter (figure 3) to test altitude and motion estimation at ETSII campus Universidad Politécnica de Madrid, Spain. The helicopter was equipped with a videre STH stereo head attached to the onboard vision computer. To run the algorithm onboard the system was fused with the helicopter attitude $(roll, pitch, yaw)$ readings and transformed to the camera coordinate system. Attitude angles transformed to camera angles serves to correct the altitude and motion estimation when the camera optical axis is not orthogonal to ground.



(a) camera looking downward (b) camera looking forward

Fig. 2. Two camera configurations for the ground robot to test a) visual odometry, b) depth estimation.



(a) helicopter side view (b) camera looking downward

Fig. 3. Camera configuration for the helicopter. Stereo camera placed on the front looking downward.

4. EXPERIMENTAL RESULT

Several indoor-outdoor experiments were performed at University of Linköping campus in Sweden. Indoor experiments were carried out to test at first glance the system inside the lab. No special landmark over a flat ground was needed for the system to work just the requirement of homogeneous ambient light. Indoor visual estimation is correlated with a predefined trajectory programmed in the robot. One trajectory was defined to test the system, a rectangular closed trajectory. Results are shown in figure 4(a). In outdoors using the mobile platform equipped with GPS, the experiments consisted in to find the correlation between the visual estimation and external referential systems like GPS. Figure 4(d) depicts the result from this experiments. This shows the correlation between the lateral-longitudinal motion estimation and GPS measurements in Easting and Northing. Figure 4(c) plots the absolute errors between vision estimation and external referential measures like GPS and robot encoder. The 2D displacement of the robot during the experiment is shown in figure 4(d). To estimate depth using stereo the robot used the configuration showed in figure 2(b) with the camera looking forward and the robot performing a longitudinal displacement of 5 m forward-backward orthogonal to a wall containing features. Vibrations and initial conditions are taken into consideration. Results are showed in figure 4(b). To evaluate the accuracy for the outdoor experiments we use two measures of the mean squared error, the error vision-GPS (MSE^V_G) and vision-robot encoder (MSE^V_E). The errors values are $MSE^V_G = 0.9470m$ and $MSE^V_E = 0.8768m$. Taking into consideration the displacement commanded of 12m the error committed is 7.83% only using visual estimation. Value that can be reduced using inertial measurements to increase robustness and accuracy.

Following, we presents the set of experiments using the COLIBRI autonomous helicopter at ETSII campus in Madrid. In the experiments the helicopter is commanded to fly autonomously following a given trajectory while the onboard stereo vision algorithm is running. The experiments find the correlation between the stereo visual estimation and the onboard helicopter state given by its sensor suite. Figures (5) and (6) shows the results from the flight trials. Four parameter are estimated for this set of experiments, the longitudinal displacement (X)[Figures (5a, 6a), lateral displacement (Y)[Figures (5b, 6b), altitude (H)[Figures (5c, 6c) and relative orientation (yaw)[Figures (5d, 6d). Altitude is computed negative since helicopter body frame is used as reference system. Each estimation is correlated with its similar value taken from the onboard helicopter state which uses a EKF to fuse onboard sensors. Table 1 shows the error analysis based on the mean square error of the visual estimation and the helicopter state. Four measures of the mean squared error are used, the error vision-GPS Northting (MSE^V_N), error vision-GPS Easting (MSE^V_E), error vision-yaw (MSE^V_ψ) and the error vision-altitude (MSE^V_H).

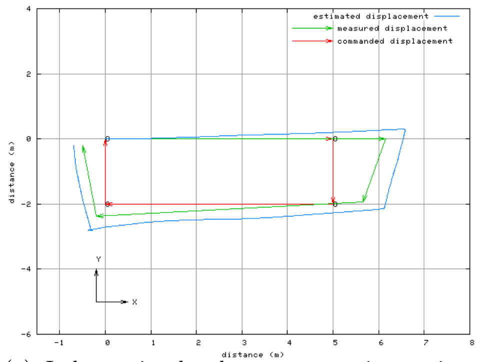
Table 1. error analysis for the helicopter experimental trials

Exp.	2	3	7	10
MSE^V_N **m**	1.0910	6.2768	1.3924	0.8531
MSE^V_E **m**	0.4712	1.8895	1.0593	1.3739
MSE^V_ψ **deg**	1.7363	8.9712	3.3288	3.0524
MSE^V_H **m**	0.1729	1.0677	1.0791	0.5077

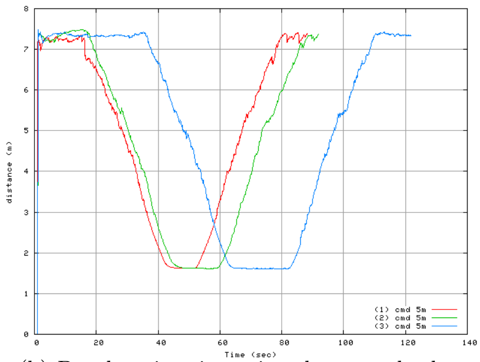
5. CONCLUSIONS AND FUTURE WORK

We have presented a stereo visual system to estimate the position and altitude of an autonomous helicopter. The developed system was tested experimentally on two different platform at two distinct locations proving its consistency. Results obtained indoors and urban outdoor scenarios allows to evaluate the performance of the approach in terms of error analysis. From the results obtained some inferences can be extracted: a) Height estimation using stereoscopy shows better robustness within the design limits. In spite of, experiments 3 and 7 show differences in same conditions error in height estimation is similar. From the experiments is observed that ECM^V_H is maintained in (13-16)% of the maximum height value. b)Motion estimation is highly dependable on the number of visually tracked features. The ICP algorithm shows mainly two lacks: inability of handle large visual displacements in the input data and susceptibility to gross outliers. c) Different factors that affect and limit the performance of system over the two platform are manifested in the results. The low translational velocity in the ground robot makes better the visual motion estimation, on the other side, vibrations and higher translational velocity in the helicopter makes the performance of the system degrades incrementing the error values. Thereafter the vision system can be used as complementary or alternative approach to GPS, incrementing the robustness where GPS coverage is transient or unavailable.

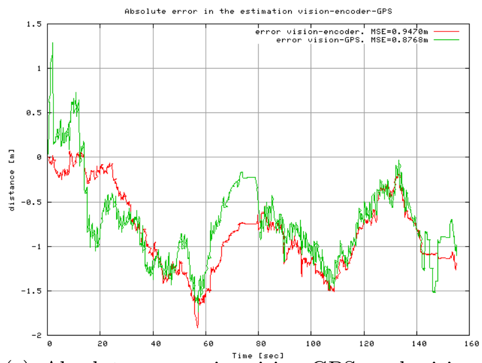
We plan to improve the capabilities of the system substituting the correlation step by a better pixel tracker algorithm. We believe this solution will help to overcome the problem of large visual displacement.



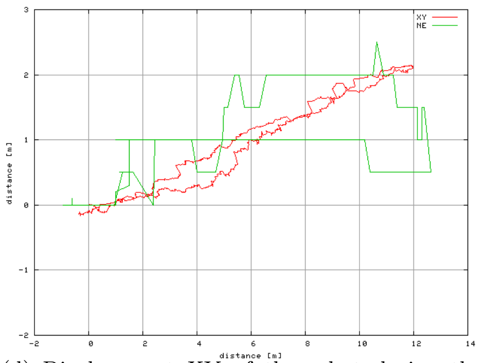
(a) Indoor visual odometry experiment in a closed trajectory.



(b) Depth estimation using the ground robot

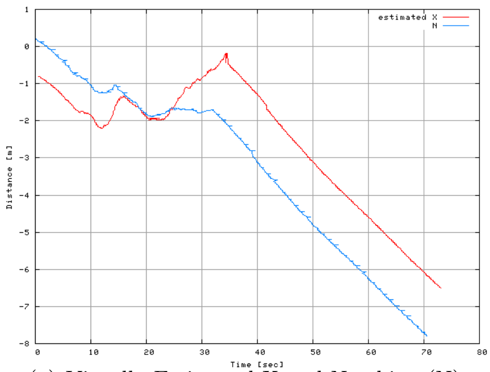


(c) Absolute errors in vision-GPS and vision-robot encoder.

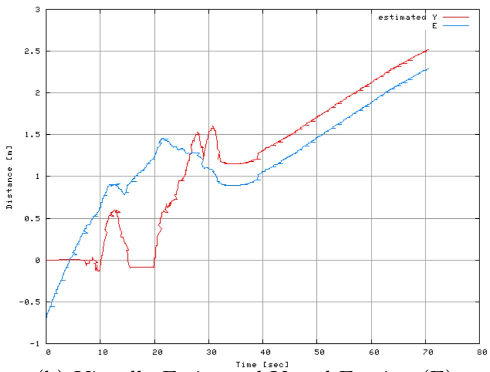


(d) Displacement XY of the robot during the outdoor experiment with GPS NE measures.

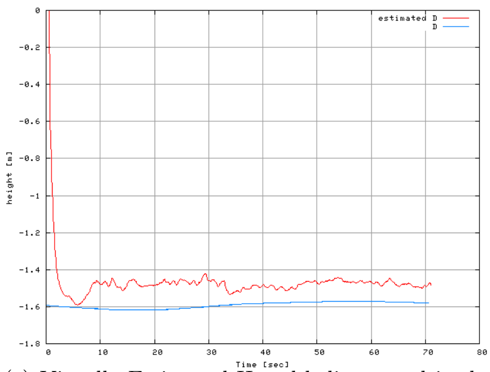
Fig. 4. Results from indoor-outdoor experiments using the ground robot.



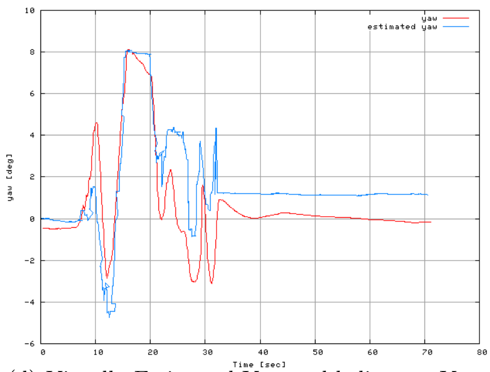
(a) Visually Estimated X and Northing (N).



(b) Visually Estimated Y and Easting (E).



(c) Visually Estimated H and helicopter altitude.

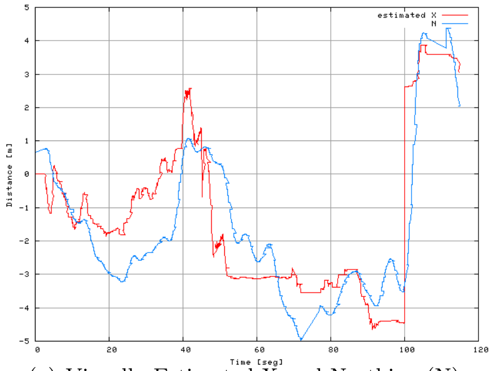


(d) Visually Estimated Yaw and helicopter Yaw.

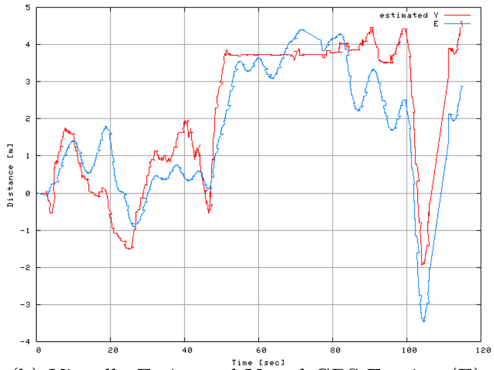
Fig. 5. Results from the second experimental trial.

6. ACKNOWLEDGMENT

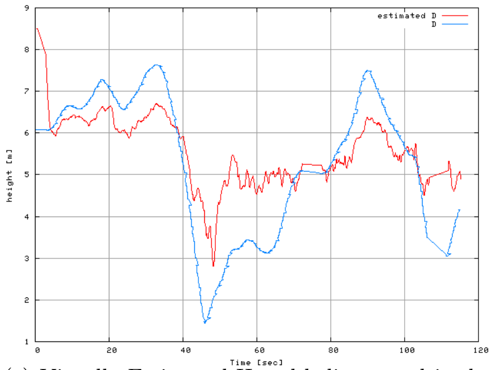
This work is sponsored in part by the Spanish Science and Technology Ministry under a project grant CICYT DPI2004-06624, and by Universidad Politécnica de Madrid-CAM. The authors would



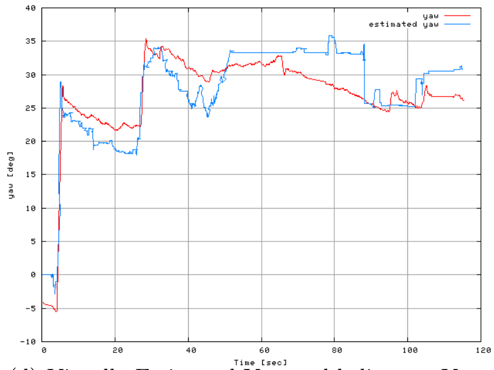
(a) Visually Estimated X and Northing (N).



(b) Visually Estimated Y and GPS Easting (E).



(c) Visually Estimated H and helicopter altitude.



(d) Visually Estimated Yaw and helicopter Yaw.

Fig. 6. Results from the seventh experimental trial.

like to thank to Jorge Leon for support with the flight trials, and Gianpaolo Conte, Piotr Rudol, Mariuz Wzorek. and the rest of the Artificial Intelligence & Integrated Computer Systems Division at LIU.

REFERENCES

Amidi, O. (1996). An Autonomous Vision-Guided Helicopter. PhD thesis. Robotics Institute, Carnegie Mellon University.

Ashbrook, A. P. (1992). Evaluations of the susan corner detection algorithm. Technical report. Department of Electronic and Electrical Engineering. University of Sheffield. Sheffield, UK.

Cheng, Yang, M. W. Maimone and Larry Matthies (2006). Visual odometry on the mars exploration rovers. *IEEE Robotics and Automation magazine* **13**(2), 54–62.

Corke, Peter, P. Sikka and Jonathan Roberts (2000). Height estimation for an autonomous helicopter. In: *Proc.Int.Symp. Experimental Robotics*. Hawaii.

Harris, C. G. and M. Stephens (1988). A combined corner and edge detection. In: *In Proceedings of the 4th Alvey Vision Conference*. pp. 147–151.

Martin, Jérome and James Crowley (1995). Experimental comparison of correlation techniques. Technical report. IMAG-LIFIA. 46 Av. Félix Viallet 38031 Grenoble. France.

Matthies, L. (1989). Dynamic stereo vision. Cmu-cs-89-195. Carnegie Mellon University. Computer Science Department.

Mejías, Luis, Ivan Mondragón, Juan Fernando Correa and Pascual Campoy (2007). Colibri: Vision-guided helicopter for surveillance and visual inspection. In: *Video Proceedings of IEEE International Conference on Robotics and Automation*. Rome, Italy.

Milella, Annalisa and Roland Siegwart (2006). Stereo-based ego-motion estimation using pixel tracking and iterative closest point. In: *Proceedings of the Fourth IEEE International Conference on Computer Vision Systems*. Washington, DC, USA. p. 21.

Moravec, H. P. (1977). Towards automatic visual obstacle avoidance. In: *Proceedings of the International Joint Cenference on Artificial Intelligence*. p. 584.

Nistér, David, Oleg Naroditsky and James Bergen (2006). Visual odometry for ground vehicle applications. *Journal Of Field Robotics* **23**(1), 3–20.

Noble, Alison (1989). Description of image surfaces. PhD thesis. Department of Engineering Science. Oxford University.