

Autonomous detection of safe landing areas for an UAV from monocular images

Sébastien Bosch and Simon Lacroix
LAAS-CNRS
Toulouse, France
Email: {firstname.name}@laas.fr

Fernando Caballero
Robotics, Computer Vision and Intelligent Control Group
University of Seville
Email: caba@cartuja.us.es

Abstract—This paper presents an approach to detect safe landing areas for a flying robot, on the basis of a sequence of monocular images. The approach does not require precise position and attitude sensors: it exploits the relations between 2D image homographies and 3D planes. The combination of a robust homography estimation and of an adaptive thresholding of correlation scores between registered images yields the update of a stochastic grid, that exhibits the horizontal planar areas perceived. This grid allows the integration of data gathered at various altitudes. Results are presented throughout the article.

I. INTRODUCTION

In the context of the development of autonomous UAVs, various functionalities that exploit environment related information are required, *e.g.* to achieve mapping tasks or mobile ground target following. Among these functionalities, the autonomous detection of safe landing areas in initially unknown environments is an important one: it can be required to ensure a safe emergency landing, or to determine areas where a payload can be delivered for instance.

If the literature abounds with contributions in which the UAV detects known landing areas (such as “H” patterns for helicopters – *e.g.* [1]), there are less contributions that deals with the detection of landing areas in unknown environments. Approaches that recover the terrain 3D geometry have been proposed to detect landing areas: in [2], inertial and accurate RTK GPS data are used to recover a digital terrain map. In [3], an approach that exploits a on-board stereovision bench is used. Others use texture and contrast information (*e.g.* [4]).

In this paper, we consider the problem of detecting safe landing areas, *i.e.* nearly horizontal obstacle-free areas, for low-cost small UAVs. This context precludes the use of heavy sensors such as a scanning laser range-finders, of a stereoscopic bench, that require a fairly large baseline to provide useful 3D over a wide range of elevations, or of heavy high quality inertial sensors. Our approach mainly relies on the *real-time* processing of monocular image sequences, and on coarse UAV attitude and motion estimates provided by low cost and light inertial and GPS sensors.

Related work

The vision literature provides sound formalisms and tools to achieve the detection of planar regions in monocular image sequences. The approaches can be classified in two main categories: the ones that explicitly recover the scene 3D structure, and the ones that do not.

In the first category, the Euclidean 3D geometry of the scene is first recovered, and then analyzed to extract the boundaries of planar regions [5], [6], [7], [8]. These approaches mainly rely on the computation of the fundamental matrix, and a batch process is often required to refine the scene reconstruction. The latter is not an option for a real-time application, and the computation of the fundamental matrix can be unstable, *e.g.* when the observed scene is mostly planar – even a stratified projective reconstruction approach fails in such cases [9]. *Plane plus parallax* approaches allow to compute both 3D parameters of the plane and the relative pose of a couple of cameras: such an approach is given in [10], but it relies on the assumption that the reference plane is parallel to the focal plane, which can in general not be guaranteed for an UAV.

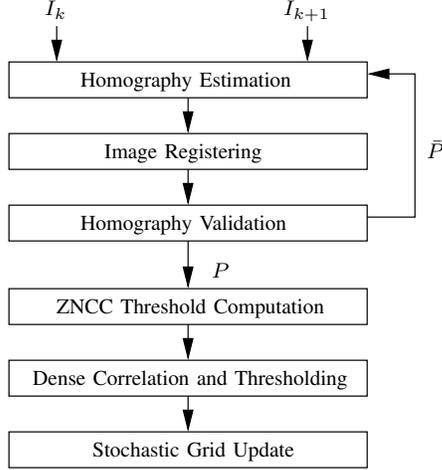
Approaches that rely on homography estimations to extract planar regions are much more suited to our context [11], [12]. They do not explicitly reconstruct the 3D geometry of the perceived scene, thus avoiding stability issues in the computation of fundamental matrices. Nevertheless, they can yield the estimation of the detected planes distance and orientation, provided coarse position and orientation information of the camera are available [13]: this can be applied to visual odometry for instance [14].

Approach and outline

Our approach to detect safe landing areas incrementally updates a model of the overflown environment that exhibits the nearly horizontal obstacle free areas, using mainly homography estimations and image correlation techniques. It can bear with mono-planar scenes scattered with obstacles or not.

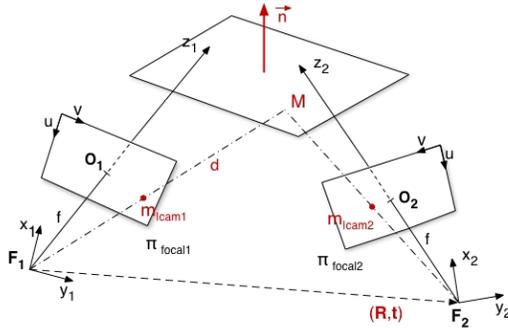
Figure 1 illustrates the processing steps applied at each image acquisition. First, an homography estimation process is applied on the basis of point matches established with the previous image. A special attention is paid to the robustness of the estimation, so as to recover a good estimate of the corresponding plane orientation (section II). The homography estimate allows to select the point matches that actually lie in the detected plane: section III presents how a dense correlation technique can be applied to determine the image areas that corresponds to the detected plane – and the ones that do not. These information are stored and managed by means of a stochastic grid updated each time the UAV gathers data, in the cells of which the probability to be planar is encoded.

Depending on the UAV altitude, the detected plane can still contain obstacles: lower altitude image acquisitions are therefore required to ensure that the detected plane is a safe landing area. Section IV presents how data gathered at different elevations are fused into a consistent representation.



1: Overview of the process to detect safe landing areas

II. HOMOGRAPHY AND PLANE ESTIMATION



2: Two views geometry of a planar scene

Consider the two views geometry represented figure 2, in which the two cameras view the same planar scene from two different positions. Let:

- $M_{|cam_i}$ the 3D point with coordinates $(X, Y, Z, 1)$ expressed in the camera i frame,
- $\bar{m}_{|cam_i}$ the central projection of M in the normalized focal plane ($z = 1m$), with coordinates $(x, y, 1)$ expressed in the camera i frame,
- $m_{|im_i}$ the projection of M in the image plane i , with pixel coordinates (u, v) in the camera i image frame.

A classical result of projective geometry [9] formulates $\bar{m}_{|cam_i}$ with respect to $M_{|cam_i}$ and $(R, t)_{1 \rightarrow 2}$, the 3D transformation between camera positions 1 and 2:

$$\begin{aligned} \bar{x}_{|cam_2} &= \frac{X}{Z}_{|cam_2} = \frac{Z_{|cam_1}}{Z_{|cam_1}} \frac{r_1 \bar{m}_{|cam_1} + t_x}{r_3 \bar{m}_{|cam_1} + t_z} \\ \bar{y}_{|cam_2} &= \frac{Y}{Z}_{|cam_2} = \frac{Z_{|cam_1}}{Z_{|cam_1}} \frac{r_2 \bar{m}_{|cam_1} + t_y}{r_3 \bar{m}_{|cam_1} + t_z} \end{aligned} \quad (1)$$

where $R = [r_1, r_2, r_3]^T$ and $t = [t_x, t_y, t_z]^T$.

Assuming that $M_{|cam_i}$ is on a plane Π defined by $ax + by + cz = d$ in the camera i frame, expression (1) yields:

$$\alpha m_{|im_2} = CNC^{-1} m_{|im_1} \quad \text{with} \quad N = \left(R + \frac{tn^T}{d} \right) \quad (2)$$

where C is the camera projection parameters (calibration matrix), $n^T = [a, b, c]$ and α is a scale factor.

$H = CNC^{-1}$ is the homography that transforms the image coordinates of any point belonging to the plane Π from the camera 1 image frame to the camera 2 image frame: the satisfaction of the relation 2 by a pair of pixels matched between images 1 and 2 states that the corresponding 3D point belong to the plane Π .

A. Estimation of H

A set of pixel matches between the two images is used to initialize the linear system:

$$Ah = 0 \quad \text{with} \quad h = [h_{11}, h_{12}, \dots, h_{ij}, \dots, h_{32}, h_{33}]^T$$

Since H is 3×3 homogeneous matrix known up to a scale factor α , we have to add an additional constraint to solve this system. Two classical solutions are possible:

- $\|h\| = 1$ which implies a $Ah = 0$ form system with h defined by 9 parameters,
- $h_{33} = 1$ which implies a $Ah = b$ form system with a defined h by 8 parameters.

In our case, the perceived planes being often parallel to the focal plane, the second solution proved to give more stable results.

B. Robust estimation

We use the algorithm proposed in [15] to establish Harris point matches between the images, and a Least Median Square estimation process [16] to eliminate the outliers, *i.e.* the matches that do not lie on the searched plane and the possible wrong matches. A Monte Carlo technique is applied to estimate the homography: assuming the outliers percentage is ϵ , the number m of draws of p matches required to have at least one draw without outliers is

$$m = \frac{\log[1 - P]}{\log[1 - (1 - \epsilon)^p]}$$

Since a Least Median Square technique does not provide a very accurate solution [17], we solve the system defined by the inliers for the found homography with a classical Singular Value Decomposition method. The whole processes is described in algorithm 1.

C. Assessing the existence of a plane

The previous algorithm allows to find an homography if the majority of the matches corresponds to a plane, but does not provide any useful information if no plane supports the majority of the matches – and also if there are no plane in the perceived scene. We therefore need a criteria to make sure that the estimated homography actually corresponds to a plane.

Algorithm 1 *Robust homography estimator*

Given an initial system $\|\mathbf{A}\mathbf{h} - \mathbf{b} = 0\|$ with $n = 2N$ entries in A .

- 1) Do uniformly m draws of p entries of A
- 2) For each draw i :
 - a) Minimize (SVD) :

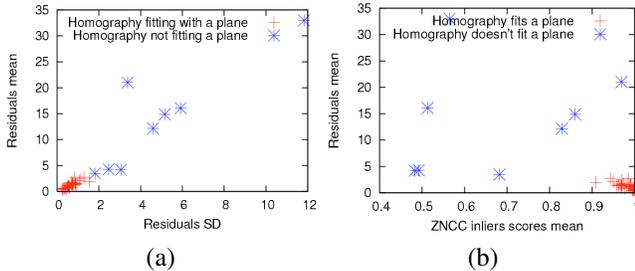
$$\hat{\mathbf{H}}_i = \underset{h}{\operatorname{argmin}} \sum_{j=1}^p [\mathbf{A}_j \mathbf{h} - \mathbf{b}_j]^2$$

- b) With $\hat{\mathbf{h}}_i$, computes the N residuals :

$$\varepsilon_{j \in [1..N]} = [\mathbf{A}_j \hat{\mathbf{h}}_i - \mathbf{b}_j]^2$$

- c) Computes the score : $\varepsilon_{med_i} = \operatorname{med}_{1..N} \varepsilon_j$
 - 3) Store $\hat{\mathbf{H}}^*$, the $\hat{\mathbf{H}}_i$ with the lowest associated ε_{med_i} .
 - 4) Keep matches fitting with $\hat{\mathbf{H}}^*$ (inliers) and resolve $\hat{\mathbf{H}}^*$ using a SVD.
-

A first idea is to analyze the repartition of the residual means and standard deviation, but an empirical analysis shows that there are no obvious thresholds than can be helpful. Figure 3-(a) shows the relations between the mean and standard deviations of the residuals for various estimated homographies, some corresponding to actual planes and some not : no discriminative threshold can be defined. But the computation of the Zero-Normalized Cross Correlation score on the matched points after the application of the homography brings a more discriminative information. Figure 3-(b) shows the repartition of the mean of the ZNCC scores and the means of the residuals: thresholds are easily defined on them to assess that the computed homography corresponds to a plane.



3: 2D projection of the three criteria used to decide if a computed homography fits with a real 3D plane

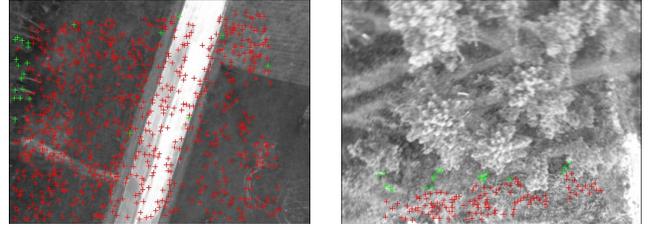
Figure 4 shows two results of the robust homography estimation process, and the identification of the matched points that belong to a plane.

D. Plane orientation estimation

Once the presence of a plane is assessed, one must make sure that its orientation is nearly horizontal to allow a safe landing. For that purpose, two solutions can be envisaged.

The first solution is to use equation 2 rewritten as:

$$n^T = \frac{d}{t^T t} t^T (C^{-1} H C - R) \quad (3)$$



4: Results of the homography estimation process based on the homography estimate. The “+” signs denote the points that have been matched : the red ones are the ones that supports an homography estimate and that have been detected as belonging to a plane. Note on the right figure that even though the matched point do not cover a wide area on the image, the process could detect the planar area.

Thanks to this relation, the orientation of the plane can be computed on the basis of the GPS and attitude measures – the elevation d is not required. However, to cope with the poor precision of the on-board sensors, this relation should be used using a rather wide baseline.

An other solution is to use a homography decomposition technique, that yields an estimation of the rotation, translation and plane normal [13]. With such an approach, two potentially correct solutions can be computed from each homography [18]: multiple views of the same scene can be used to select the correct one.

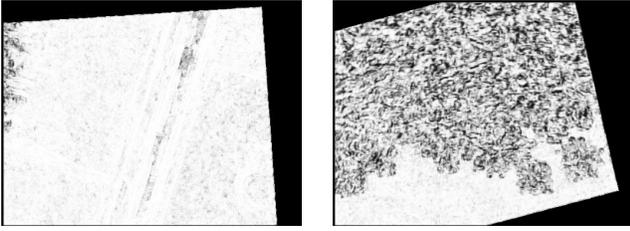
III. PLANAR SURFACE DETECTION ALGORITHM

Up to now, we know that the matched points retained for the homography estimate do correspond to a nearly planar surface, but this information is rather useless to assess the areas where the UAV can land. This section describes how the image regions that correspond to the plane are extracted: first, the acquired images are registered with respect to the first image in which an horizontal plane has been detected. An analysis of the stabilized image correlation segments the image pixels into planar and non-planar areas, and a probabilistic grid structure is updated.

A. Dense plane detection

Once an homography that corresponds to a horizontal plane has been estimated, its application to the current image generates an image that can be registered with the reference image. Assuming a sufficient parallax effect (see section IV-A), non-planar image regions should be badly correlated, whereas the pixels that do lie on the estimated plane will match in the two registered images (figure 5): as a consequence the registered image correlation scores give useful information on the image regions that correspond to the detected plane.

An important issue is to determine which image in the sequence should be correlated with the reference image: one one hand, one would like to have the distance between the camera positions as large as possible, so as to estimate the plane parameters with a good precision. Also, a wider baseline allows to detect smaller obstacles (see section IV-A). On the other hand, the overlap between the images should be large enough to contain a large number of point matches, thus yielding a precise homography estimate. We currently use the following approach:



5: ZNCC scores computed between the registered images whose matches are shown in figure 4 – the whiter the pixels, the higher the correlation score and the probability that they correspond to the detected plane.

- 1) A first image is chosen as the reference frame,
- 2) For each newly acquired image, the interest points matching algorithm is run, and the existence of an homography that corresponds to a horizontal plane is checked,
- 3) If a horizontal plane is detected *and* the overlap area with the reference images is below 70% of the image surface, the current image is registered and correlated with the reference image, and the probabilistic grid structure is updated,
- 4) The last registered image is taken as a new reference and the whole process is re-iterated.

B. Registered images analysis

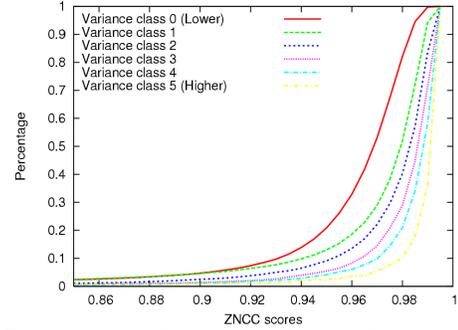
Experimental trials show that it is not possible to determine a static value for a threshold on the ZNCC scores that would yield a good segmentation in planar and non planar areas of the images. Indeed, despite the normalization properties of ZNCC, important score variations are observed, depending on the scene texture and illumination, and on the camera properties (dynamic of the camera, sharpness of the images...).

We propose here a to adapt the threshold on the ZNCC score to the variance of the considered pixels. We made statistics over hundreds of registered images, and recorded the ZNCC scores of the matched interest points that do correspond to a plane. Figure 6 shows the scores cumulative distribution function for each variance class : as reported by [19] it clearly appears that the scores corresponding to low variance pixels are usually lower than the ones corresponding to higher variance pixels. These observations lead us to determine the automatic threshold algorithm 2.

Algorithm 2 Automatic ZNCC score thresholding

- 1) Classify the pixel variances of the interest points that supports the homography.
 - 2) Classify ZNCC scores of inliers points with respect to variance class C_{σ_i} .
 - 3) Get the median value med_i of each ZNCC score class C_{znci} (*i.e.* the score that separates the cumulative distribution function into two equal areas)
 - 4) Compute the threshold τ_i for each score class C_{σ_i} as $\tau_i = 1 - 3 * (1 - med_i)$
-

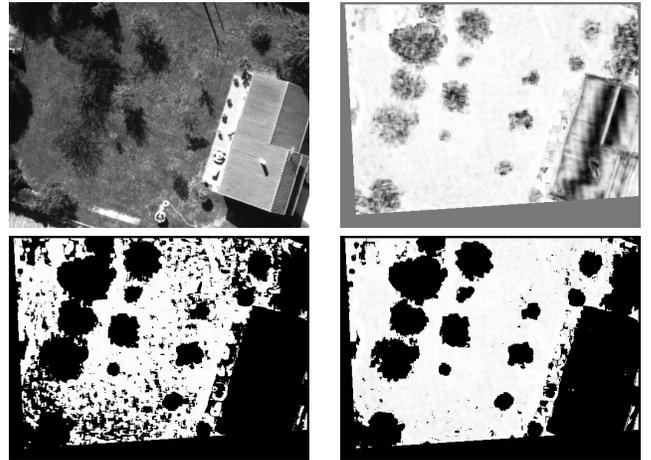
The definition of $\tau_i = 1 - 3 * (1 - med_i)$ has been empirically defined, and has shown to yield good separation



6: ZNCC scores cumulative distribution function for 6 pixel variances classes.

results. The use of the median value to define this threshold is much more stable than the minimal value determined for each score class C_{σ_i} . Indeed, there are at most a few hundreds matched points for two registered images: even though they are inliers, their small number provides a poor score sampling.

Once the automatic thresholds are defined on the basis of the matched points that correspond to the detected plane, they are used to separate all the registered images pixels in two classes. Figure 7 shows a result of this thresholding process.



7: Illustration of the thresholding process to determine the area corresponding to the detected horizontal plane. (a): reference image. (b): computed ZNCC scores between the registered images. (c): simple thresholding based on a threshold defined as $\mu_{znc} + k\sigma_{znc}$: many areas corresponding to the plane have been removed. (d): result of our variance adaptive thresholding algorithm: all the obstacles are detected (no non detection), and only a few pixels corresponding to the plane have been labeled as obstacle.

C. Fusing informations into a grid structure

Now that we are able to segment from two registered images the areas that do correspond to a planar surface, we store this information in a data structure, so as to be able to incrementally fuse the informations gathered as the UAV is flying.

We use a grid model, that is referenced with respect to the first image for which a plane had been detected, and that has a resolution specified as k times the image resolution (one grid cell is covered by $k \times k$ pixels). Each cell of the grid is

updated with a probability of being planar every time a new registered image includes it.

The computation of the cell's probabilities relies on the classic Bayes rule, using the probabilities of good detection P_d and of false detection P_f :

- P_d is the ratio between pixels labeled as planar and the ones that are actually planar.
- P_f is the ratio between pixels labeled as planar and that are not.

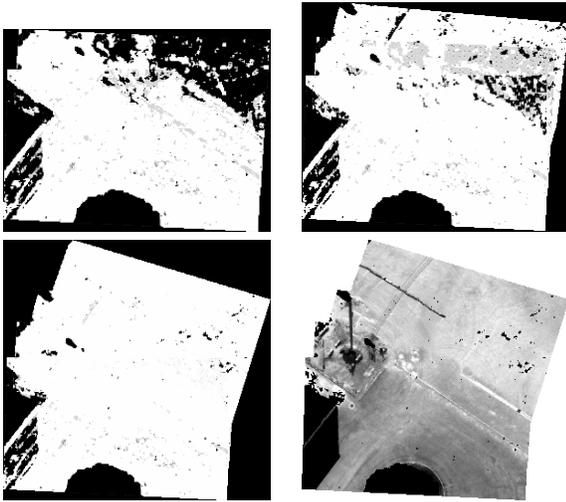
Empirical statistics on various illustrative images have lead us to adopt the values $P_d = 80\%$ and $P_f = 12\%$.

At the initialization of the grid structure, the probability of each cell to be planar is set to $p(0) = 1/2$. As new observations are made, the cell probabilities are updated according to the following formulas:

$$p(k+1) = \frac{P_d p(k)}{P_d p(k) + P_f (1-p(k))}$$

if the pixel fused in the cell is planar, and otherwise :

$$p(k+1) = \frac{(1-P_d)p(k)}{(1-P_d)p(k) + (1-P_f)(1-p(k))}$$

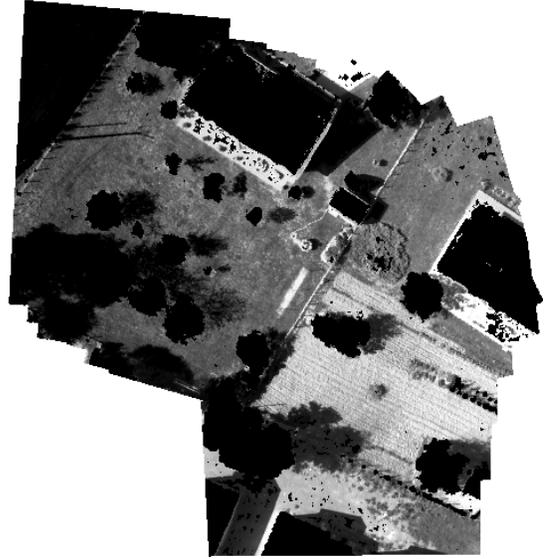


8: Probabilistic grid refinement process with three images. The bottom right image shows the reconstructed mosaic with only the pixels that have been labeled as being planar.

Figure (8) shows the refinement process using three images updates. While the grid structure is updated, we also update a mosaic of the pixels that are detected as being planar: this mosaic is required to register grids structure obtained from various passes over the terrain (section IV). The grey values of the mosaic are normalized in order to compensate the luminosity variation between the images (figure 9).

Performance: On a usual laptop, the overall process depicted in figure 1 and the interest point detection and matching step can be applied at a rate of about $2Hz$ on 512×384 images.

Table I shows the computed normals of the plane shown figure 9, using the homography decomposition approach mentioned in section II-D: the various normal estimates are very similar, and show that the approach allows a quite good



9: Mosaic of a planar surface reconstructed from a sequence of 40 images – the black areas have been identified as non-planar.

detection of the plane orientation. Note however that these normals are expressed in the reference frame of the first image, but a coarse attitude information suffices to asses that the plane is nearly horizontal.

Image	n_x	n_y	n_z
5	0.1493	0.1436	0.9783
6	0.1436	0.0909	0.9855
11	0.1093	-0.0045	0.9940
15	0.0619	0.0159	0.9980
20	0.0740	-0.0562	0.9957

I: Some normals computed for the plane detected in figure 9, using various homography estimates. The Z axis is perpendicular to the camera in the outward direction

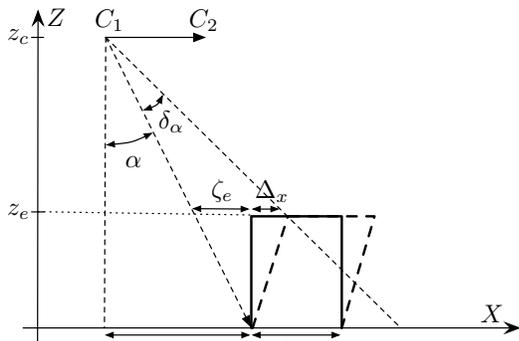
IV. GRID REFINEMENT

A. Sensitivity of the approach

The magnitude of the parallax is of course the determinant factor that specifies the height of the obstacles that can be detected. We led a theoretical analysis to analyze the relations between the elevation, the baseline, the camera resolution and field of view, in order to establish the values of the height of an obstacle over a plane that produces a one pixel shift when registering the image with the homography corresponding to the plane (figure 10). Table II summarizes the obtained results.

	$z_c = 30m$	$z_c = 50m$	$z_c = 100m$
$z_e(\tau_r = 30\%, \alpha = 0^\circ)$	0.0428m	0.0714m	0.143m
$z_e(\tau_r = 50\%, \alpha = 0^\circ)$	0.06m	0.1m	0.2m
$z_e(\tau_r = 50\%, \alpha = 15^\circ)$	0.0356m	0.0593m	0.118m
$z_e(\tau_r = 50\%, \alpha = 30^\circ)$	0.0242m	0.0406m	0.0807m
$z_e(\tau_r = 70\%, \alpha = 0^\circ)$	0.1m	0.166m	0.333m

II: Elevation z_e of an obstacle that yields a 1 pixel shift in the registered images, as a function of the camera altitude z_c , orientation α and overlapping rate τ_r that implicitly defines the baseline (see notations figure 10 – the camera characteristics chosen to establish these figures are the ones of the camera we used to produce the results shown throughout the paper)

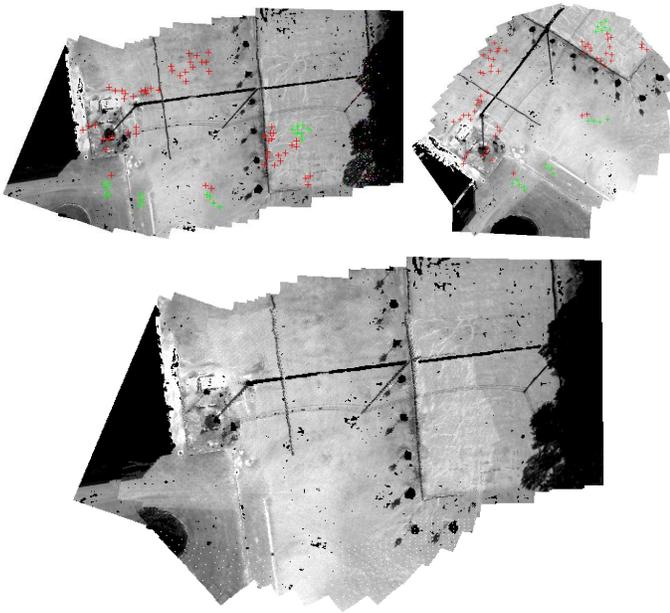


10: Characterization of the detectable elevation after the image registration.

B. Fusing grids acquired over several passes

It is clear that the approach could benefit from the integration of data acquired at various altitudes. A realistic scenario of the exploitation of our method would be to first fly over the terrain at a rather high altitude, select in the grid the areas that appear as planar, and then to perform a flight at a lower altitude in order to detect smaller obstacles. Also, the integration of several passes at similar altitude could help to assess large planar areas (e.g. to detect landing areas for fixed-wing UAVs).

The possibility to fuse the various probabilistic grids built is therefore relevant in various contexts. However, the grids are not geo-referenced, because we do not consider that the UAV is equipped with a precise GPS and especially a precise attitude sensor. But the mosaic built along with the grids can be used to register the grids, using the Harris interest point matching algorithm. Once registered, the cells probabilities can be updated thanks to the classical Bayes formula. Figure (11) illustrates the fusion of two grids.



11: Registration of the mosaics associated to two grids built during two different passes. The two images on top show the initial grids, with the matched interest points used to register them. The bottom image is the resulting mosaic.

V. SUMMARY AND FUTURE WORKS

We have presented an approach that allow the autonomous detection of safe landing areas detection for an UAV, that do not require any precise localization and orientation on-board sensor. On the basis of a robust homography estimation technique and of an automatic determination of the threshold to apply on the correlated registered images, the approach is able to extract the dominant plane in the perceived scene. The results are stored in a probabilistic grid data structure, that allows the fusion of several passes over the terrain.

A further analysis of the way to update the grid cell probabilities needs to be conducted. In particular, the probabilities of good and false detections P_d and P_f associated to the labeled pixels should take into account the variance of the pixels. Similarly, the influence of the baseline between the registered images should be considered.

Besides its integration on-board an actual UAV, we are also considering other applications of this approach: detection of moving elements on the ground, and air/ground robot cooperation. In the latter case, the approach can readily be used to detect navigable areas for a ground rover.

REFERENCES

- [1] S. Saripalli, J. Montgomery, and G. Sukhatme, "Vision-based autonomous landing of an unmanned aerial vehicle," in *IEEE ICRA*, 2002, pp. 2799–2804.
- [2] A. Johnson, J. Montgomery, and L. Matthies, "Vision guided landing of an autonomous helicopter in hazardous terrain," in *IEEE ICRA*, 2004.
- [3] M. Meingast, C. Geyer, and S. Sastry, "Vision based terrain recovery for landing unmanned aerial vehicles," in *IEEE Conference on Decision and Control*, 2004.
- [4] P. Garcia-Padro, G. Sukhatme, and J. Montgomery, "Towards vision-based safe landing for an autonomous helicopter," *Robotics and Autonomous Systems*, vol. 38, no. 1, pp. 19–29, 2001.
- [5] C. Baillard and A. Zimmerman, "Automatic reconstruction of piecewise planar models from multiple views," in *Proc. of the CCVPR*, 1999.
- [6] A. Bartoli, "Piecewise planar segmentation for automatic scene modeling," in *Proc. of the IEEE ICCVPR*, vol. II, 2001, pp. 283–289.
- [7] H. Oriot and A. Michel, "Building extraction from stereoscopic aerial images," *Applied Optics*, vol. 43, pp. 218–226, 2004.
- [8] M. Lourakis, A. Argyros, and S. Orphanoudakis, "Detecting planes in an uncalibrated image pair," in *BMVC02*, 2002, p. Poster Session.
- [9] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.
- [10] R. Dick, AR. Cipolla, "Model refinement from planar parallax," in *10th British Machine Vision Conference*, 1999, pp. 73–82.
- [11] E. Vincent and R. Laganière, "Detecting planar homographies in an image pair," in *IEEE International Symposium on Image and Signal Processing and Analysis*, Croatia, 2001, pp. 182–187.
- [12] Y. Kanazawa and H. Kawakami, "Detection of planar regions with uncalibrated stereo using distributions of feature points," in *The 15th British Machine Vision Conference*. September, 2004, pp. 247–256.
- [13] B. Triggs, "Autocalibration from planar scenes," in *Proc. of ECCV*, 1998.
- [14] F. Caballero, L. Merino, J. Ferruz, and A. Ollero, "A visual odometer without 3d reconstruction for aerial vehicles. applications to building inspection," in *Proc. of the ICRA*, 2005, pp. 4684–4689.
- [15] I.-K. Jung and S. Lacroix, "A robust interest point matching algorithm," *International Conference on Computer Vision*, July 2001.
- [16] P. Rousseeuw, "Least median of squares regression," *American Statistics Association*, vol. 79, pp. 871–880, 1984.
- [17] T. Hettmansperger and S. Sheather, "A cautionary note on the method of least median squares," *American Statistician*, 1992.
- [18] S. Saripalli and G. Sukhatme, "Landing on a mobile target using an autonomous helicopter," in *Proceedings of the International Conference on Field and Service Robotics, FSR*, July 2003.
- [19] H. Oriot, "Extraction de bâtiments de formes complexes à partir de couples d'images aériennes stéréoscopiques." Onera, Tech. Rep. RT 1/06761 DTIM, 2003.