Real-Time Reconstruction of Heightmaps from Images Taken with an UAV

Jose Gabriel Ramirez-Torres and Ander Larranaga-Cepeda

Abstract In this paper, an approach for a real-time three dimensional reconstruction of a dense point cloud, from images obtained with an unmanned aerial vehicle, is presented. The approach also estimates the model scale, using a known marker as reference, obtaining a fairly precise heightmap of the environment. The proposed image analysis framework is also capable of an off-line 3D reconstruction from monocular images, without known markers in the environment. The proposed framework is an adaptation of the Patch-based Multiview Stereo (PMVS) algorithm, in order to obtain a faster point cloud generation, suitable for real-time environment reconstruction. For the obtained heightmap, the error estimation ranges between 1–1.5 % for manual detection of marker and 4–5 % for automatic marker detection, which we consider is accurate enough for autonomous navigation and path planning.

Keywords Real-time 3D reconstruction · UAV applications · Image analysis · Computer vision · Autonomous robots

1 Introduction

An unmanned aerial vehicle (UAV) can be defined as a powered aerial vehicle that does not carry a human operator, uses aerodynamic forces to provide vehicle lift and can fly autonomously or be remotely piloted [9]. UAVs are classified according

J.G. Ramirez-Torres (✉) · A. Larranaga-Cepeda
Laboratory of Information Technologies, CINVESTAV Tamaulipas, Tamaulipas, Mexico
e-mail: grtorres@tamps.cinvestav.mx

A. Larranaga-Cepeda
e-mail: alarranaga@tamps.cinvestav.mx

to the type of wing (fixed or rotary wing) and to its taking off capabilities (vertical take-off or short take-off and landing) [4]. Good performance, low cost, vertical/short take-off and landing allowing access to hazardous areas for humans, low weight and low noise, can be counted among the main advantages of UAVs for different applications. At present, semi-autonomous UAVs are just able to perform very specific automatic tasks, such as marker-based navigation, automatic take-off and landing [20]. In order to expand the actual range of possible applications of UAVs, we must assure a safe and autonomous navigation through unknown environments, which requires a robust mapping ability of the surroundings. Such exploring ability leads to many non trivial problems, such as the UAV localization problem in a GPS denied area or the real-time obstacle evasion when navigating through an unknown environment. In order to solve these problems, it is required to generate a dependable metric representation of the environment, based on the received data from sensors, that is, a map containing distance, height and angle data of the overflown area.

The most common way to describe the orography of the environment is through heightmap. The generation of a heightmap is a process involving the digital representation of the terrain surface from a sequence of sensor readings (rangefinder or visual sensors), in order to obtain a numeric model known as Digital Elevation Model (DEM), Digital Terrain Model (DTM) or Digital Surface Model (DSM) [22, 25]. The main applications of digital terrain models are the terrestrial parameters extraction, generation of three-dimensional maps for cartographic relief depiction, orthorectification of aerial photographies, analysis of GPS data, analysis of terrain surfaces, flight planning, among others.

In litterature, there exist different proposals for digital terrain model generation from GPS-based images obtained using a UAV [2, 3, 20]. The purpose of this paper is to propose a solution based on a commercial UAV equipped with a monocular camera, in order to provide an easy and affordable way for the generation of digital terrain models, particularly in a GPS-denied environment, useful for academic and civilian purposes. For this work, the selected UAV model is an AR.Drone 1.0, manufactured by the French company Parrot; however, the presented development can be easily adapted to any UAV, with a good quality camera for image acquisition.

This work proposes a digital image analysis for the generation of a point cloud, both online and offline, in order to obtain a three dimensional model with scale estimation, as well as the heightmap construction, exploitable for UAV autonomous navigation. The paper is organized as follows: Sect. 2 describes the previous work in three dimensional modeling, heightmap generation and the odometry and navigation algorithms for the AR.Drone UAV; Sect. 3 describes the structure of the proposed approach for the image analysis; Sect. 4 present the obtained experimental results; Sect. 5 provides a data analysis and interpretation; and finally, Sect. 6 summarizes and presents some final conclusions.

2 Previous Work

In this section, some relevant works on image-based 3D reconstruction and heightmaps generation are presented.

2.1 Image-Based 3D Reconstruction

Some approaches, addressing the three dimensional reconstruction of the environment from monocular images, have been published. Due to the complexity of this task, many authors suggest the use of multiple sensors for this purpose [19] or stereoscopic cameras [27] in order to obtain depth data from the images and discard the monocular cameras as not suitable for 3D reconstruction [19]. There are also some other approaches focused on Visual SLAM using an Extended Kalman Filter (EKF) for the three dimensional reconstruction of facades, but with scale uncertainty [6, 16]. For the approaches based on structure-from-motion (SfM) with a single camera, the scene can be recovered, but only up to a scale, since the absolute dimensions cannot be computed [21]. To solve the scale uncertainty, the presence of a known reference in the scene must be assured, so the absolute scale can be recovered from the relative dimensions of the reference in it. We propose to employ a marker that could be easily detected by an algorithm for marker detection, such as the method proposed by Garrido-Jurado et al. [8].

The offline 3D reconstruction problem from images has been successfully addressed with algorithms like the Patch-based Multiview Stereo algorithm (PMVS) [7] and the SfM approaches, such as A Contrario Structure from Motion algorithm (ACSfM) [17], capable to provide a dense and highly detailed point cloud. However, the required processing time for image matching and camera pose computing is too high for real-time three dimensional reconstruction and, on the other hand, the scale factor is also unknown.

2.2 Scale Estimation for Heightmaps

There exist several proposals when it comes to heightmap generation from aerial images, but most of them are based on GPS data [1, 26] and stereoscopic cameras [2] to get depth data. Weiss's et al. [26] approach creates a three dimensional model in real time after manually measuring the scale factor, however, only environments with available GPS information are considered. Also, several robust feature descriptors have been proposed for defining an object in a three dimensional space [5, 15, 28], but most of them do not consider occlusions and cannot deal properly with several instances of the same object in a scene.

3 Proposed Approach

Simply put, the UAV, equipped with a monocular camera, must overfly an indoor area or a GPS-denied terrain, while the proposed system analyses the images and constructs a heightmap of the environment.

To define the path that the UAV must follow, we use the marker method proposed by Garrido-Jurato et al. [8] for setting navigation checkpoints on the terrain, assigning instructions for the UAV to each marker. The marker provides three additional and useful advantages: first, the matrix of the camera pose can be estimated in real-time, with relatively little computational cost overhead; second, the a priori knowledge of the size of the marker allows us to find out the point cloud's scale factor; and third, the marker also sets the ground level of the scene, that is, the XY plane used as reference for all the estimated heights.

For the point cloud generation, a real-time image analysis, based on the PMVS algorithm [7], is used. The points of interest (POI) are the corners and the edges of the image, obtained by the Harris' algorithm [10] and the difference of gaussians (DoG) [14] respectively. Each POI is modeled as a patch and can appear in more than one image. When a new image is analysed, patches which are epipolar consistent are merged, according to a photometric consistency criterium measured by cross-correlation. The process works as follows: the images are analysed by blocks of three images and a temporary point cloud is obtained, which is used to update the general point cloud. In this way, the point cloud grows as the UAV advances through the environment.

In order to precisely detect the marker within the scene, we use an adaptation of the Tombari's et al. [24] approach for object detection in a three dimensional space, based in the voting procedure of the Hough Transform [11], adapted for 3D shapes. In our approach, there are two point clouds to be compared: the actual modeled environment, called *scene*, and the point cloud describing the marker, called *model*. The total number of points in both clouds are counted and their normal vectors are computed according to Khoshelman's method et al. [12]. Then, a uniform sampling is performed on both clouds in order to reduce them into small squares and approximating every square by its centroid.

Afterward, the reduced clouds are described by SHOT descriptors, as depicted in [23] and, by applying Euclidean distance, matching pairs among descriptors are found. If a matching is found, the BOARD algorithm [18] assigns a local reference frame (LRF) in order to make it rotational and translational invariant. Finally, Hough voting is performed to create an histogram of orientations with all local reference frames. A variable threshold defines the number of found point correspondences, so it can be concluded whether the reference object (*model*) is present within the *scene* or not. If the marker has been found, a search for its corners is performed to measure it. The relative size in the point cloud and the absolute size of the marker allow to solve the scale uncertainly and obtain a real metric 3D model of the environment.

The proposed approach has been implemented in C/C ++ on a Linux OS, structured in two parts: the first for the control of the UAV and the reconstruction of the temporary point cloud, the second for the global point cloud updating and the marker searching. The complete schema of the proposed system is presented in Fig. 1.

As an additional feature, the system is capable of generating a 3D model from the images, without a known marker, in an off-line procedure. This task is accomplished with the ACSfM algorithm [17] which performs image matching based on the SIFT algorithm [13] in order to compute the camera pose matrices for each image. Finally, with the computed pose matrices, the whole set of images is processed with the PMVS algoritm to generate the global point cloud of the environment.

4 Results

To validate the proposed approach, two distinct scenarios, with different sizes of marker were used, (Fig. 2). For both scenarios, the resulting heightmaps are shown in Fig. 6. Four known heights *A*, *B*, *C* and *D* were used as reference for computing
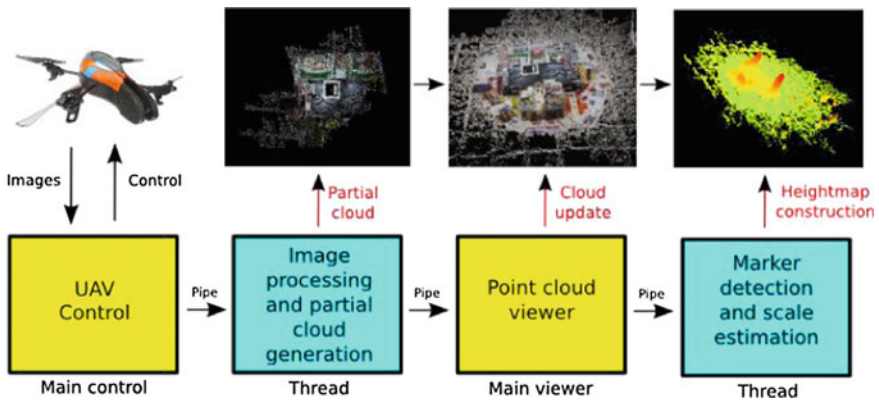

Fig. 1 Proposed approach for real-time 3D heightmap reconstruction

Fig. 2 Scenarios 1 and 2

the estimation error (c.f. Fig. 2). A 14 × 14 cm maker was used in Scenario 1 and a 21 × 21 cm maker was used in Scenario 2 to evaluate the estimation error for different marker sizes.

4.1 Real-Time 3D Modeling

A set containing 72 640 × 480 images was used to obtain the model shown in Fig. 3. The growing process of the global point cloud is shown in Fig. 4. The Fig. 5 shows the relief of the 3D model.

4.2 Marker Detection and Heightmap Generation

Tables 1 and 2 show the estimated height and the error for real heights *A*, *B*, *C* and *D* for each scenario respectively, with different threshold values for Hough voting (Fig. 6).

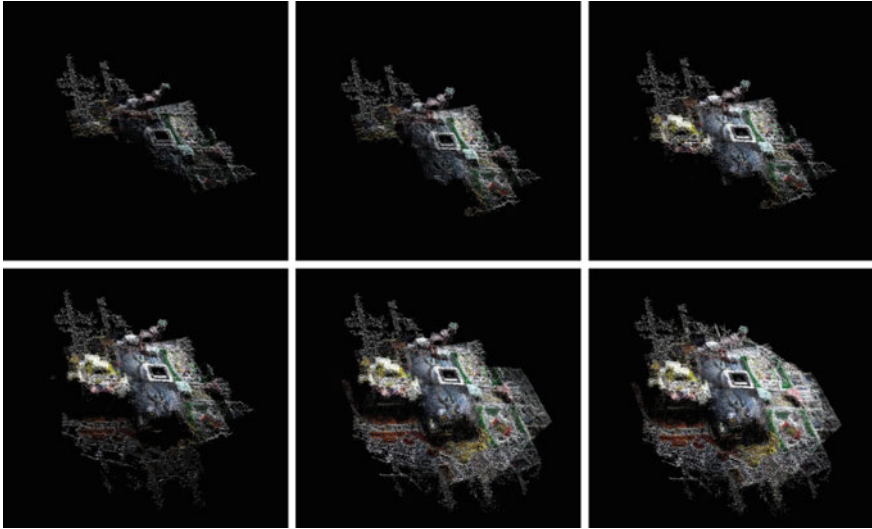**Fig. 3** Example of the image data set used for the presented results
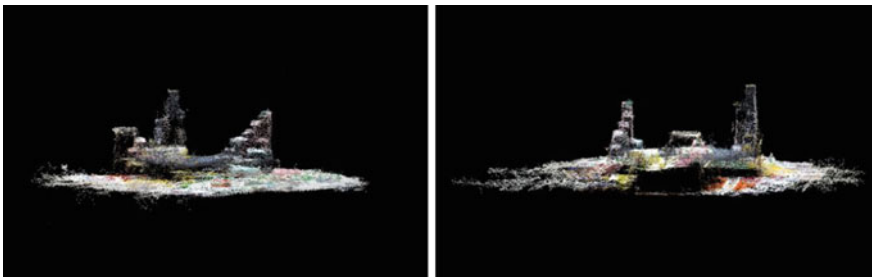
Fig. 4 Real time reconstruction of the point cloud

Fig. 5 3D profile of the generated model

Table 1 Height estimation results for Scenario 1, with 14 cm marker

A (0.235 m)	B (0.39 m)	C (0.16 m)	D (−0.14 m)	Avg error (%)	Threshold
0.2308	0.4046	0.1719	−0.1369	3.79	0.01
0.2645	0.4477	0.1952	−0.1523	14.53	0.015
0.3176	0.5474	0.2310	−0.1962	40.01	0.020

Table 2 Height estimation results for Scenario 2, with 21 cm marker

A (0.39 m)	B (0.395 m)	C (0.55 m)	D (0.24 m)	Avg error (%)	Threshold
0.4195	0.4098	0.5645	0.2471	4.22	0.01
0.4261	0.4039	0.5707	0.2457	4.41	0.015
0.4329	0.4143	0.5814	0.2525	6.70	0.020
0.4398	0.4181	0.5777	0.2513	7.09	0.030
0.5171	0.4985	0.7001	0.3071	28.51	0.1
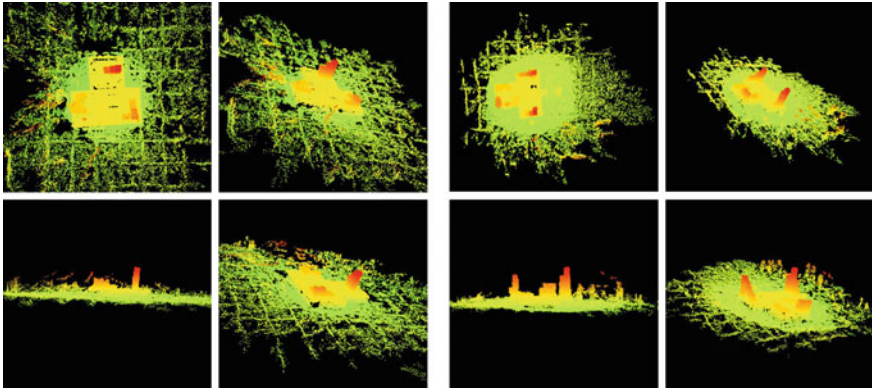
Fig. 6 Generated heightmap for Scenario 1 and 2, respectively

4.3   *Off-line 3D Modeling*

The off-line 3D model generated from the same 72 640 × 480 images is shown in Fig. 7.

5   Discussion

The results reported in Tables 1 and 2 are slightly different because not only the size of the marker, but also the threshold in Hough voting affects the scale estimation process. The image quality and resolution have a significant influence on the camera pose estimation and on the marker detection in the point cloud. This lead us to conclude that scale estimation relies on an accurate marker detection and, at the same time, the precision for the camera pose matrix relies on the visual quality of the marker. For autonomous navigation, 640 × 480 images provide a heightmap in an adequate processing time, enough to accurate fulfill the established requirements of the present work. In these scenarios, a processing time of 1.05 frames per second on average was reached.
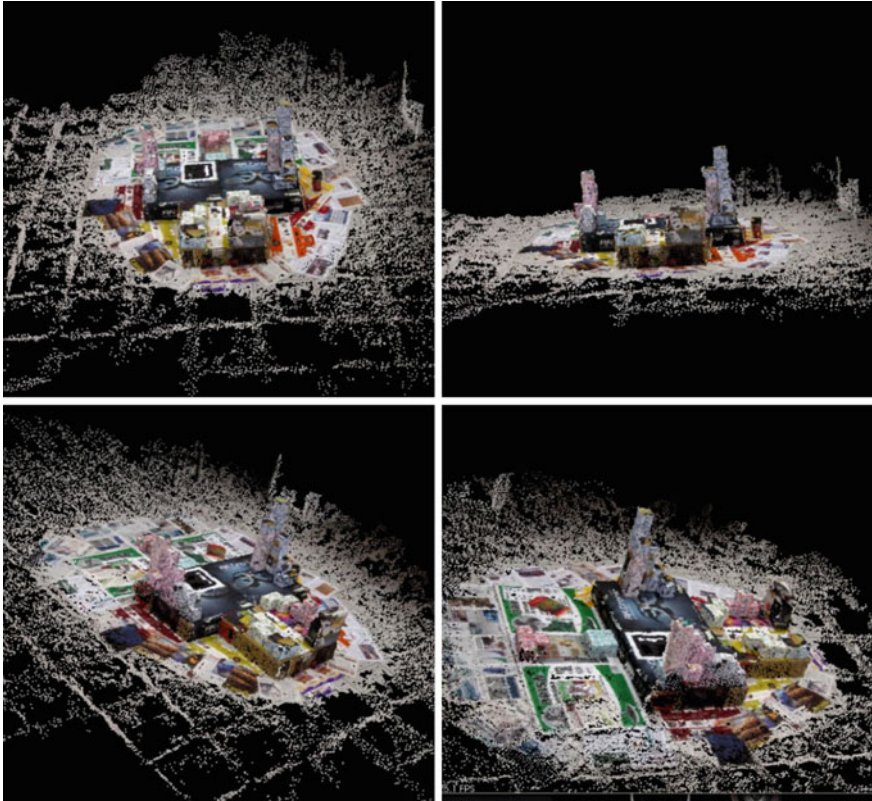
Fig. 7 Different perspectives of the obtained off-line 3D model

The combination of a SfM algorithm for computing camera poses and a patch-based point cloud generation algorithm provided an accurate, dense and highly detailed 3D model in the off-line approach. For this case, the image set is treated after the UAV has overflown the environment.

6   Conclusion

Based on the obtained results, it can be concluded that the generated 3D model is comparable and even better, in terms of visual quality and processing time, when compared to the previous works presented in Sect. 2, for both real-time and off-line approaches, but including the global scale factor estimation. Also, the height estimation errors are suitable and accurate enough for a safe autonomous navigation through the environment. It is planned as future work to test the presented framework with real aerial images to study how the environment and the altitude of

the UAV affects the heightmap generation. Processing time reduction, heightmap construction without artificial markers, UAV autonomous indoor navigation and collaborative real-time 3D modeling are also considered.

References

1. Barazzetti, L., Scaioni, M.: Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation. Photogram. Rec. **25**, 2010 (2013)
2. Besnerais, G., Sanfourche, M., Champagnat, F.: Dense height map estimation from oblique aerial image sequences. Comput. Vis. Image Underst. **109**, 204–225 (2008)
3. Call, B.: Obstacle avoidance for unmanned air vehicles. Master's thesis, Brigham Young University (2006)
4. Cetinsoy, E., Dikyar, S., Hancer, C., Oner, K., Sirimoglu, E., Unel, M., Aksit, M.: Design and construction of a novel quad tilt-wing UAV. Mechatronics **22**, 723–745 (2012)
5. Chen, H., Bhanu, B.: 3D free-form object recognition in range images using local surface patches. Pattern Recogn. Lett. **28**, 1252–1262 (2007)
6. Diskin, Y., Asari., V.: Dense point-cloud creation using superresolution for a monocular 3D reconstruction system. In: Proceedings of SPIE 8399, Visual Information Processing XXI, pp. 83,990–N–83,990N–9 (2012)
7. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. Pattern Anal. Mach. Intell. **32**(8), 1362–1376 (2010)
8. Garrido-Jurado, S., Muoz-Salinas, R., Madrid-Cuevas, F., Marn-Jimnez, M.: Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recogn. **47**(6), 2280–2292 (2014)
9. Gertler, J.: US unmanned aerial systems. Congressional Research Service (2012)
10. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference, pp. 147–151 (1988)
11. Hough, P.: Method and means for recognizing complex patterns (1962)
12. Khoshelham, K.: Extending generalized hough transform to detect 3D objects in laser range data. In: Proceedings of ISPRS Workshop on Laser Scanning, pp. 206–210 (2007)
13. Lowe, D.: Distinctive image features from scale-invariant key points. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
14. Marr, D., Hildreth, E.: Theory of edge detection. In: Proceedings of the Royal Society of London, pp. 215–217 (1980)
15. Mian, A., Bennamoun, M., Owens, R.: On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes. Int. J. Comput. Vis. **89**, 348–361 (2010)
16. Motta, C.: Reconstruccin tridimensional de fachadas de edificios empleando imgenes monoculares obtenidas por un vehculo areo no tripulado autnomo. Master's thesis, Centro de Investigacin y de Estudios Avanzados del Instituto Politcnico Nacional (2014)
17. Moulon, P., Monasse, P., Marlet, R.: Adaptive structure from motion with a contrario model estimation. In: ACCV 2012, vol. 7727, pp. 257–270 (2012)
18. Petrelli, A., Stefano, L.D.: On the repeatability of the local reference frame for partial shape matching. In: 13th International Conference on Computer Vision (ICCV), pp. 2244–2251 (2011)

19. Pollefeys, M., Nistér, D., et al.: Detailed real-time urban 3D reconstruction from video. Int. J. Comput. Vis. **78**, 143–167 (2008)
20. Sanfourche, M., Delaune, J., Besnerais, G.L.: Perception for UAV: vision-based navigation and environment modeling. J. AerospaceLab **4**, 1–19 (2012)
21. Scaramuzza, D., Fraundorfer, F., Pollefeys, M., Siegwart, R.: Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 1413–1419 (2009)
22. Sirmacek, B., d'Angelo, P., Krauss, T., Reinartz, P.: Enhancing urban digital elevation models using automated computer vision techniques. In: International Conference on Pattern Recognition (2010)
23. Tombari, F., Salti, S., DiStefano, L.: Unique signatures of histograms for local surface description. In: ECCV'10 Proceedings of the 11th European Conference on Computer Vision Conference on Computer Vision, pp. 356–369 (2010)
24. Tombari, F., Stefano, L.D.: Object recognition in 3D scenes with occlusions and clutter by hough voting. In: 4th Pacific-Rim Symposium on Image and Video Technology, pp. 349–355 (2010)
25. Weibel, R., Heller, M.: Digital terrain modelling (1993)
26. Weiss, S., Achtelik, M., Kneip, L., Scaramuzza, D., Siegwart, R.: Intuitive 3D maps for MAV terrain exploration and obstacle avoidance. J. Intell. Robot. Syst. **61**(1–4), 473–493 (2011). Doi:10.1007/s10846-010-9491-y, http://dx.doi.org/10.1007/s10846-010-9491-y
27. Wen-Chung, C., Shu-An, L.: Real-time feature-based 3D map reconstruction for stereo visual guidance and control of mobile robots in indoor environments. In: 2004 IEEE International Conference on Systems, Man and Cybernetics, vol. 6, pp. 5386–5391 (2008)
28. Zhong, Y.: Intrinsic shape signatures: a shape descriptor for 3D object recognition. In: Proceedings of 3DRR Workshop (in conj. with ICCV), pp. 689–696 (2009)