

# Building and site reconstruction from small scale unmanned aerial vehicles (UAV's)

Friedrich Fraundorfer  
Institute for Computer Graphics and Vision  
Graz University of Technology, Austria  
Email: fraundorfer@icg.tugraz.at

**Abstract**—This paper investigates the use of small scale unmanned aerial vehicles (UAV's) for the 3D reconstruction of buildings and urban sites. The paper discusses the current state-of-the-art and benefits of using UAV imagery. Furthermore, the paper describes the steps of a highly automated 3D reconstruction pipeline. Experiments are carried out to show the benefits of UAV imagery for urban building reconstruction and results are compared to these of traditional aerial imaging.

## I. INTRODUCTION

Recent advances in robotics have led to the development of small scale unmanned aerial vehicles (UAVs) [1]. One particular interesting design of such flying robots is the multi-rotor helicopter platform which has the capability of on-the-spot hovering as compared to fixed-wing designs. The possibility of GPS-based autonomous flight and the possibility of equipping UAVs with different types of sensors (e.g. digital cameras) make these systems highly interesting for remote sensing applications, in particular as alternatives to conventional airborne systems. One specific field of application that will benefit from these developments is that of 3D map generation in urban environments. Due to the small size and weight, UAVs can carry sensors very close to urban structures and can move into positions that provide unobstructed views onto the surveyed structures, thus being able to create measurements that would not be possible with conventional airborne sensor systems. 3D city models generated from airborne images typically lack fidelity, e.g. low image resolution at the facades of buildings. Image acquisition and processing typically use nadir imagery. But even for oblique imagery the view to facades could be blocked by nearby buildings or vegetation such that a complete 3D modeling of buildings in urban environments is infeasible. This is where UAV imagery could be used to fill in these missing data. Fig. 1 illustrates this idea. Fig. 1(a) shows a nadir image taken from an airborne imaging system (Ultracam). The facade of the building is barely visible, 3D modeling with sufficient resolution is not possible from such an image. Fig. 1(b) shows an image taken from a camera mounted on a UAV. The UAV can move into positions where the view to the facade is unobstructed and thus can deliver high resolution images of facades of buildings. The challenge is now to fuse these two types of image data together to create one complete model of a building. In particular for processing image data taken by UAVs new methods have to be developed as the standard tools for aerial 3D processing are inadequate for typical UAV imagery. For automatic building reconstruction utilizing UAVs we propose the following processing pipeline:



(a)



(b)

Fig. 1. Comparison between standard aerial image quality and UAV image quality. (a) Standard aerial image. Facades of buildings are typically not visible. (b) Image taken from a UAV. Facades and roof areas can be imaged with the same sensor at high resolution.

(a) automatic data capture, (b) 3D image registration, (c) 3D processing, (d) geo-referencing and (e) 3D data fusion. Automation of all of the individual steps is crucial and new algorithms still need to be developed to achieve fully automatic processing without human interaction.

## II. STATE-OF-THE-ART

The state-of-the-art in urban modeling and 3D reconstruction is represented by traditional aerial imaging or satellite imaging, e.g. for large scale mapping [2], [3], [4] or urban monitoring [5]. Although these methods achieve impressively high ground resolution (which is still increasing) they provide limited view to e.g. the facades of buildings. Image data acquisition using UAV's can provide the necessary views to

complement traditional aerial or satellite data, however, so far UAV's have only been used to reconstruct isolated buildings or sites [6], [7], [8], [9], [10] and have not been employed for large scale mapping. The listed works have in common that UAV's are used to capture images of an isolated site to produce a 3D reconstruction. In most cases this is already a fully automated process. However, fusing these isolated 3D reconstruction into DSM's or DEM's from large scale mapping is labor intensive manual work. The main challenge is automatic geo-referencing of the UAV 3D reconstructions. Sensors for direct geo-referencing typically don't fit on small scale UAV's which means that geo-referencing methods have to make use of the captured image data, e.g. by using ground control points [11]. To automate this process and also to automate the next step, 3D data fusion, are challenges that still remain.

### III. PROPERTIES OF UAV IMAGE DATA

The qualities of UAV image data differ from qualities of standard airborne imagery. Manned, large aircrafts can carry heavy high resolution digital cameras (e.g. 260MPixel Ultracam). Small scale UAV's of a type that can fly close to buildings are able to carry cameras with a weight of up to 1-2kg. This limits camera resolution currently to about 20MPixel. When interested in high-frame rates the available camera resolution is again much lower, e.g. 1-2MPixel. In standard airborne image acquisition data is acquired by flying along a regular grid. UAV's that operate close to the ground typically can't maintain such a grid-like flight trajectory. The necessity to avoid obstacles and the necessity to achieve unobstructed viewing positions will lead to irregular flight trajectories which makes image registration a difficult task. Image registration cannot make use of accurate position information from additional sensors e.g. INS system as these systems are too large to be fitted on small scale UAV's. This means, that image registration has to be done by image processing relying on the image content only.

### IV. PROCESSING PIPELINE FOR UAV IMAGE DATA

#### A. Automatic data capture

For GPS-controlled UAV's data capture can be fully automatic. Using a mission planning software (e.g. QGroundControl<sup>1</sup>) a flight plan can be defined and the UAV is following this plan from start to landing without an operator. However, when doing the planning an operator has to make sure, that the flight plan does not collide with existing structure, e.g. buildings or trees. To make even the process of flight planning automatic we investigated the idea of autonomous exploration. In previous work [12] we utilized on-board processing to detect and avoid obstacles in-flight based on camera images. Ultimately, this will result in a fully robotic image acquisition system.

#### B. 3D image registration

In the image registration step, 6DOF camera poses have to be computed for every image in the acquired data set. This is a very challenging task for UAV image data as typically no accurate camera pose estimation is available from

direct measurements (e.g. as in contrast to the UltraCam camera with integrated INS/GPS system in the form of the UltraNAV system). To solve this problem we developed a software package that is tailored to be used with UAV imagery, the MAVMAP software [6]. The key steps of this software are automatic wide baseline image matching, outlier-robust camera pose estimation and 3D feature point triangulation. The last step is a bundle adjustment robust to still remaining outliers from mis-matches. Automatic feature matching, in particular matching of wide baseline images, is in almost all cases producing outliers from mis-matches. All the pose estimation algorithms need to be robust against such outliers. In MAVMAP this is achieved using robust estimators, e.g. RANSAC and robust cost functions. In addition to robust algorithms, additional constraints from data of an inertial measurement unit (IMU) are also used. Basically all UAV's are equipped with accelerometers and gyroscope that form an IMU for flight control. These systems, due to scale and weight limits, cannot be used to track the full pose of the UAV but it is possible to measure the orientation of the UAV for each image shot. This orientation can be used as additional constraint in pose computation and makes the optimization more robust against outliers. These constraints can be used for pose optimization by minimizing the cost function Eq. 1.

$$E_r = \sum_{k=1}^n \rho_x \left( \| \mathbf{x}_k - q(\mathbf{P}_i, \mathbf{X}_j) \|_2^2 \right) + \lambda \sum_{l=1}^m \rho_r \left( \left\| \hat{\mathbf{R}}_l - \mathbf{R}_l \right\|_F \right) \quad (1)$$

Eq. 1 is minimized by varying the camera poses  $P_i$  and the 3D points  $X_j$ , with  $x_k$  being the image measurements,  $q$  being the image projection function and  $\rho_x$  and  $\rho_r$  being Cauchy-functions as robustifier.  $\hat{R}_l$  are the measured rotations from the IMU and  $R_l$  are the rotational parts of the camera poses  $P_i$ , with  $\lambda$  being a weight term between image measurements and IMU measurements.

#### C. 3D processing

3D processing is the step of computing 3D points from image measurements by means of dense stereo matching. For any two or more images a 3D point cloud can be computed and registered into a common coordinate system using the previously computed camera poses. The accuracy of the registered point clouds depends on the accuracy of the previous camera registration step. For dense stereo matching elaborate methods like semi-global matching [13], multi-view stereo [14] can be used.

#### D. Geo-referencing

In the geo-referencing step the 3D reconstruction from UAV images is transformed into a geo-referenced coordinate system. Standard method is to add unique markers to the scene or to use available man-made structures (e.g. man-holes) and then to do a manual geodetic measurement of these ground control points (GCP) to obtain the necessary coordinate transformation. Another method is to align the GPS positions of the UAV with the camera poses from the sparse reconstruction. However, the accuracy of the GPS receivers on a moving UAV is low leading to a much bigger uncertainty

<sup>1</sup><http://www.qgroundcontrol.org/>

in geo-referencing than compared to a method that uses GCP. For this step, new fully automatic methods yet need to be developed. One promising approach would be to match the new UAV imagery with already geo-referenced aerial image data. However, different image resolution, different acquisition times and different viewing angles pose huge difficulties for automatic processing.

### E. 3D data fusion

In the simplest case 3D data fusion can mean to merge the point clouds from traditional aerial imaging with point clouds from UAV imaging to complement the 3D reconstruction. However, duplicated measurements of the same structures with different uncertainties will lead to unwanted artifacts. To deal with this problem a more sophisticated fusion algorithm will be necessary. Possible solutions are grid based fusion methods or volumetric fusion methods [15], but current approaches have severe limitations in resolution and scene size.

## V. EXPERIMENTS

We perform experiments for the different steps of the previously described pipeline on the **InfUAV** data set. The data set was captured by using the Asctec Falcon UAV system, equipped with a 10MPixel camera. The data set consists of 8 images of a building with facade and parts of the roof area visible (see Fig. 1(b) for one image of this data set). For comparison to standard aerial imaging we created the **InfUltra** data set taken by an UltraCam digital aerial camera. Each UltraCam image has a resolution of 130MPixel.

### A. 3D image registration of UAV image data

Our MAVMAP software was used to automatically compute the camera poses and initial sparse 3D points from the **InfUAV** data set. A rendering of the camera poses and the point cloud can be seen in Fig. 2. The final average re-projection error is 0.3px. From the rendering it can be seen that the positions from which the images are taken are at irregularly spaced positions, a typical property of UAV image data sets.

### B. 3D processing of UAV data

In this experiment we analyze the results of 3D processing of an image pair from the **InfUAV** data set. The cameras have been pre-registered using MAVMAP as explained in the previous section. The 3D processing step of MAVMAP has then been used to do dense stereo 3D processing to compute a depth-map based on two images and to finally compute a 3D point cloud. A rendering of this point cloud can be seen in Fig. 3. The point cloud of this part of the facade consist of 250000 depth measurements. In contrast, this facade imaged by the UltraCam aerial camera (**InfUltra** data set) amounts to only 25000 pixels total, only a tenth of the resolution (see Fig. 1 for the actual image of the facade). In principle it would also be possible to utilize the full resolution of the images of the **InfUAV** data set to achieve 3.7 million depth measurements for this specific facade. Table I summarizes these numbers.

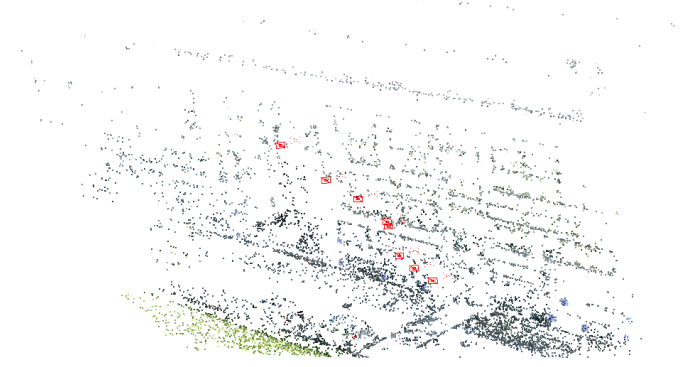


Fig. 2. A rendering of the camera poses and sparse point cloud obtained by MAVMAP for the **InfUAV** data set.

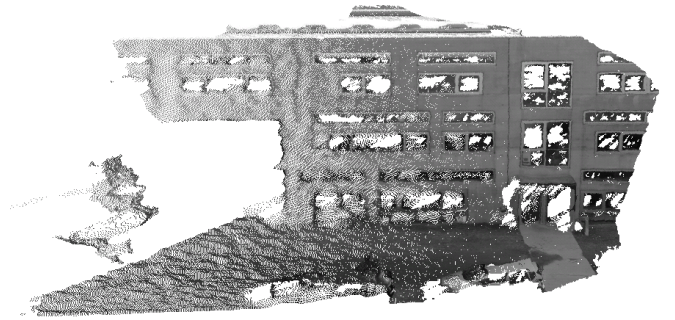


Fig. 3. Rendering of the dense 3D point cloud computed from the **InfUAV** UAV image data set.

TABLE I. COMPARISON OF THE IMAGING RESOLUTION FOR THE SAME BUILDING FACADE WHEN USING STANDARD AERIAL IMAGERY AND UAV IMAGERY. FIG. 3 SHOWS A 3D POINT CLOUD OF THE FACADE COMPUTED FROM THE UAV IMAGERY.

data set	facade resolution in pixel
UAV imagery ( <b>InfUAV</b> )	3.7 million
Aerial imagery ( <b>InfUltra</b> )	25000

### C. Geo-referencing and data fusion

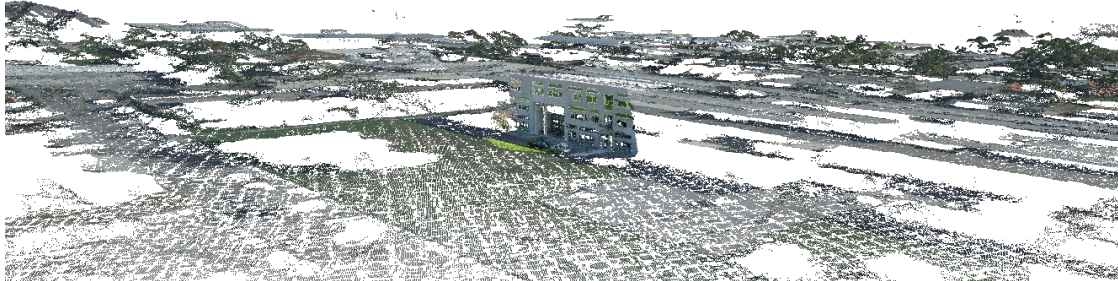
In this experiment the point cloud computed from the **InfUAV** data set gets geo-referenced by matching to ground control points specified in the **InfUltra** data set. To this end, matching ground control points in both data sets are defined manually. The extreme view point change and change in resolution does not allow for automatic image matching. After geo-referencing the point cloud of the **InfUAV** data set it is merged into the point cloud of the **InfUltra** data set. No sophisticated fusion process was used for this, the point cloud data was just added. Fig.4 shows a rendering of the merged point cloud. Fig.4(a) shows the whole point cloud in which the part from the **InfUAV** is colored in red. From this rendering it is visible that the point cloud from the **InfUltra** data set does not contain building facades. Fig.4(b) shows a cut out with higher details. One can see that the building facade gets added to the point cloud from the **InfUAV** data set.

## VI. CONCLUSION

This paper described the current state-of-the-art in building and site reconstruction from small scale UAV systems and



(a)



(b)

Fig. 4. Rendering of merged point clouds from both data sets **InfUltra** and **InfUAV**. (a) Overview rendering. The UAV point cloud is colored in red. (b) Cut out to see the details of the **InfUAV** point cloud. The UAV imagery fills in the details of the facade in high quality.

discussed its challenges. The paper presented the necessary steps for processing UAV image data and gave details about the individual steps. Experimentally the benefits and advantages of UAV imagery have been compared to traditional aerial imaging. It has been demonstrated that UAV imagery can complement traditional aerial imagery to make more complete urban 3D reconstructions.

#### ACKNOWLEDGMENT

The author thanks the Aerial Vision Group at Graz University of Technology for providing the **InfUAV** and the **InfUltra** data sets.

#### REFERENCES

- [1] D. Scaramuzza, M. Achtelik, L. Doitsidis, F. Fraundorfer, E. Kosmatopoulos, A. Martinelli, M. Achtelik, M. Chli, S. Chatzichristofis, L. Kneip, D. Gurdan, L. Heng, G. Lee, S. Lynen, L. Meier, M. Pollefeys, A. Renzaglia, R. Siegwart, J. Stumpf, P. Tanskanen, C. Troiani, and S. Weiss, "Vision-controlled micro flying robots: from system design to autonomous navigation and mapping in gps-denied environments," *IEEE Robotics and Automation Magazine*, vol. 21, no. 3, 2014.
- [2] A. Wiechert, M. Gruber, and K. Karner, "Ultramap: The all in one photogrammetric solution," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXIX-B3, pp. 183–186, 2012.
- [3] F. Kurz, S. Türmer, O. Meynberg, D. Rosenbaum, H. Runge, P. Reinartz, and J. Leitloff, "Low-cost optical camera systems for real-time mapping applications," *PFG Photogrammetrie, Fernerkundung, Geoinformation*, vol. 2012, no. 2, pp. 159–176, 05 2012.
- [4] P. d'Angelo, M. Lehner, T. Krauss, D. Hoja, and P. Reinartz, "Towards automated dem generation from high resolution stereo satellite images," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Conference*, 2008, pp. 1137–1342.
- [5] H. Chaabouni-Chouayakh, P. d'Angelo, T. Krauss, and P. Reinartz, "Automatic urban area monitoring using digital surface models and shape features," in *Urban Remote Sensing Event (JURSE), 2011 Joint*, April 2011, pp. 85–88.
- [6] J. L. Schönberger, F. Fraundorfer, and J.-M. Frahm, "Structure-from-motion for mav image sequence analysis with photogrammetric applications," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-3, pp. 305–312, 2014.
- [7] A. Gruen, Z. Zhang, and H. Eisenbeiss, "Uav photogrammetry in remote areas - 3d modeling of drapham dzong bhutan," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXIX-B1, pp. 375–379, 2012.
- [8] M. Cramer, S. Bovet, M. Gültlinger, E. Honkavaara, A. McGill, M. Rijsdijk, M. Tabor, and V. Tournadre, "On the use of rpas in national mapping - the eurosdr point of view," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W2, pp. 93–99, 2013.
- [9] P. Fallavollita, M. Balsi, S. Esposito, M. G. Melis, M. Milanese, and L. Zappino, "Uas for archaeology - new perspectives on aerial documentation," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W2, pp. 131–135, 2013.
- [10] J. Bartelsen, H. Mayer, H. Hirschmüller, A. Kuhn, and M. Michelini, "Orientation and dense reconstruction of unordered terrestrial and aerial wide baseline image sets," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 1-3, pp. 25–30, 2012.
- [11] F. Bachmann, R. Herbst, R. Gebbers, and V. V. Hafner, "Micro uav based georeferenced orthophoto generation in vis + nir for precision agriculture," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W2, pp. 11–16, 2013.
- [12] L. Heng, D. Honegger, G. H. Lee, L. Meier, P. Tanskanen, F. Fraundorfer, and M. Pollefeys, "Autonomous visual mapping and exploration with a micro aerial vehicle," *Journal of Field Robotics*, vol. 31, no. 4, pp. 654–675, 2014.
- [13] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 2, pp. 328–341, Feb 2008.
- [14] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, June 2006, pp. 519–528.
- [15] G. Vogiatzis, P. Torr, and R. Cipolla, "Multi-view stereo via volumetric graph-cuts," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, June 2005, pp. 391–398 vol. 2.