FULLY AUTOMATIC UAV IMAGE-BASED SENSOR ORIENTATION

L. Barazzetti [a], F. Remondino [b], M. Scaioni [a], R. Brumana [a]

[a] Politecnico di Milano, Dept. B.E.S.T., p.za L. da Vinci 32, 20133 Milan, Italy
(luigi.barazzetti; raffaella.brumana; marco.scaioni)@polimi.it
[b] 3D Optical Metrology Unit, Bruno Kessler Foundation (FBK), Trento, Italy
Web: http://3dom.fbk.eu, E-mail: remondino@fbk.eu

Commission I, ICWG I/V

KEY WORDS: Automation, Bundle Adjustment, Image orientation, Unmanned Aerial Vehicles

ABSTRACT:
The paper presents a procedure for the automatic orientation of image blocks captured by UAV-borne cameras. Disregarding the availability or absence of GPS/INS data, the method is able to extract a set of manifold tie points to be used as observations in a photogrammetric bundle adjustment. Complex image configurations and scenes can be dealt with, encompassing the use of convergent imagery, strong perspective deformations, lighting changes and so on. Tie points are extracted among image pair combinations by exploiting operators like SIFT and SURF, which are invariant with respect to scale variations and rotations as well. Then all image pairs are linked together to compute a global bundle adjustment. In order to speed up the process, a coarse-to-fine approach can be used, integrated to the creation of a *visibility map* that limits the selection of image pairs to be analysed.
Some tests on different UAV blocks are reported and discussed, showing the potential of this procedure to obtain the orientation of all images in a fully automated way.

1. INTRODUCTION

In recent years, UAVs demonstrated their great potential for photogrammetric measurements in many application fields. UAV photogrammetry (Colomina *et al.*, 2008) indeed opens various new applications in the close range domain and introduces also a low-cost alternatives to the classical manned aerial photogrammetry (Eisenbeiss, 2009). In particular, rotary or fixed wing UAVs, capable of performing the photogrammetric data acquisition with amateur or SLR digital cameras, can fly in manual, semi-automated and automated flight modes and are generally coupled with GPS/INS sensors. The presence of GPS/INS onboard is usually exploited to guide the acquisition of a regular block and, theoretically, it could allow the direct georeferencing of the captured images. But bundle adjustment is generally computed, starting from the approximated exterior orientation (EO) parameters provided by the integrated GPS/INS, to further refine the sought accurate camera poses and attitudes. In the literature several examples using this method are reported (Eisenbeiss, 2008; Grenzdörffer *et al.*, 2008). In other applications, where a low-metric quality is required, e.g. for fast data acquisition and mapping during emergency response (Zhou, 2009), the accuracy of GPS/INS can be enough.
On the other hand, there are many surveying applications and scenes where the objects to be surveyed or their location are not suitable (even partially) for the use of direct georeferencing techniques. These difficulties are due to (i) a prevalent vertical extension of the target object (like a big building façade, see Pueschel *et al.*, 2008; Scaioni *et al.*, 2009), which needs the acquisition of several convergent or wide baseline images close to the object, or (ii) the necessity to operate in environments where satellite visibility is limited or impossible at all (downtowns, rainforest areas, etc.). In both cases the GPS positioning can be hardly used even for autonomous flight modes as the signal is strongly degraded. Therefore, an image-based approach would be the only viable way to perform the image orientation phase (Eugster and Nebiker, 2008; Wang *et al.*, 2008).

In addition, the small size and the reduced pay-load of most UAV platforms limit the transportation of high quality IMU devices like those coupled to airborne cameras or LiDAR sensors. GPS is mainly used in code-based positioning mode and it does not give a sufficient precision for direct sensor orientation, nonetheless when space- or ground-based augmentation systems (Hofmann-Wellenhof, 2008), nor DGPS positioning are employed. The use of RTK techniques would improve the quality of positioning to a centimetre level, but the system would become too complex and heavy.
For the orientation of large image blocks, the interactive tie points extraction might require a long elaboration time, thus automation becomes necessary. Likewise in aerial photogrammetry, where the *automatic aerial triangulation* (AAT – Büyüksalih and Li, 2005) is today a standard consolidated technique, the UAV image orientation procedure should work in autonomous way. The presence of scale differences, illumination changes, occlusions and convergent imagery results in a failure of standard AAT with commercial software, as these problems are much closer to those usually afforded in close-range applications (Luhmann *et al.*, 2006). On the other hand, the deployment of targets can help for the automation of the image orientation task, but in many applications where UAVs are used, this operation is too much complex, time consuming and costly to be accomplished.
Following these concerns, we present a fast and accurate methodology to automatically extract image correspondences from UAV image blocks and then to derive the EO parameters through a bundle adjustment. The method is derived from a procedure developed for the automated orientation of different types of terrestrial data sets, like generic image sequences, sparse blocks, sequences of multispectral imagery and spherical images. Other recent publications of the authors report more technical details of the orientation algorithms and several successful examples (see Barazzetti *et al.*, 2009; 2010). The automated method is based on the integration of Computer

Vision and photogrammetric methods to extract a great number of well distributed tie points to obtain the EO of a block captured with one or more pre-calibrated digital cameras. Comparing to existing commercial approaches, no manual intervention is required (besides the identification of the GCPs). Robust techniques are employed to find sets of accurate and sub-pixel correspondences between the images. The whole procedure can be considered as a multi-step process in which several parameters are estimated and refined, whilst the EO is continuously improved through the computation of a standard photogrammetric bundle adjustment. The automated process usually provides very huge sets of tie points, according to the object texture and the image block geometry. Nevertheless, when the size of the block is large (more than 25-30 images), such tie point abundance might result in computational problems within the bundle adjustment numerical solution. To overcome this drawback, a method for tie point decimation based on their local redundancy and regular geometric distribution in the images has been implemented (Barazzetti *et al.*, 2010).

The automated tie points extraction and image orientation procedure were tested using different UAV datasets and four salient examples are reported in the following sections. Accuracy analyses with ground truth data achieved with standard manual measurements are also presented.

2. THE IMAGE ORIENTATION PROCEDURE

The implemented procedure for image orientation (ATiPE - *Automatic Tie Point Extraction*) is based on the preliminary extraction and matching of scale invariant features and the following bundle solution. The method was developed to orient close-range image blocks and adapted to images taken with UAVs. In fact, the existing commercial softwares used in aerial photogrammetry can only use automatic algorithms to extract image correspondences and determine the EO parameters of the images, starting from GPS/INS data. However, the sensors mounted on UAVs generally do not have a high accuracy and are mainly used for the autonomous navigation of the vehicle. Moreover, the use of convergent images complicates the automatic identification of the image correspondences. For these reasons we implemented an *ad hoc* algorithm capable of working without approximated EO parameters, although these data can be employed.

The flowchart of the algorithm is shown in Figure 1. The input elements are the images and their calibration parameters. An image compression can be employed to speed up the processing with a coarse-to-fine approach. A conversion of each original image to a new greyscale one is performed by extracting the green channel. On the other hand, other combinations of channels could be used in the case other sensors are used.

The block is then divided into image pairs considering all possible image combinations (n images give $(n^2-n)/2$ pairs) and for each possible combination points are extracted and matched with the SIFT (Lowe, 2004) or SURF (Bay *et al.*, 2008) operators using the *ratio test* on the L2 norm of the descriptors (Snavely *et al.*, 2006). As the images generally provide a good number of features, a restrictive threshold is employed in this phase. The images are normally elaborated at their original size, providing a huge number of homologous points. To speed up the search inside each pair an approximate procedure based on a *kd-tree* is used (Arya *et al.*, 1988) and outliers are removed with the robust estimation of the relative orientation concept encrypted in the *fundamental* or *essential matrices* (Hartley and Zisserman, 2004). As robust estimator, the *MAPSAC* method is

employed (Torr, 2002). The entire procedure has a quadratic computational cost with respect to the number of images. This means that in the case of large blocks in which images are processed at their original size, the CPU time becomes an issue of essential importance. To reduce the number of image pairs a procedure was developed in order to detect the image combinations sharing tie points. This method creates the so called *visibility map* by using two different automated procedures:

- *visibility from images*: if high-resolution images are employed, a preliminary elaboration with compressed images (e.g. less than 2 Mpx) is rapidly performed. This derives the image combinations of the whole block. Then, the same procedure is repeated with the original images but taking into account the produced image combination map;
- *visibility from GPS/INS data:* these values combined with an approximate DTM or DSM of the area allow the estimation of the overlap among the images. This method is faster than the previous one. However, it can only be applied for images with a configuration similar to that of an aerial block. In some cases, the DTM can be approximated with a plane.
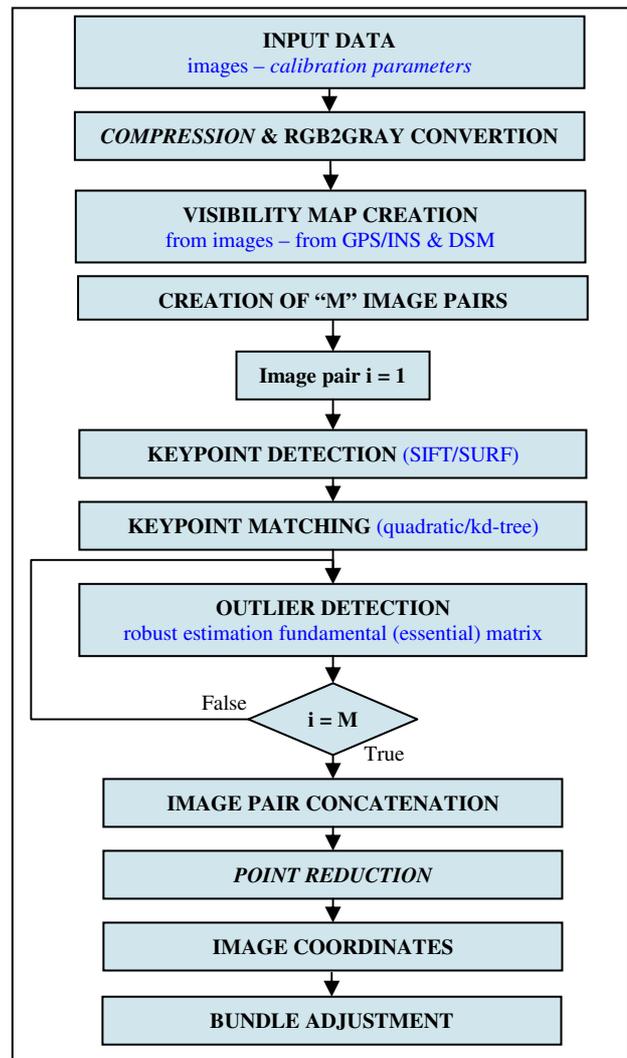
Figure 1. The flowchart of the implemented method for automatic tie point extraction and successive bundle adjustment. In *italic* the optional data or functions.

After the matching of all possible image pairs, these are combined using the numerical values of the detected image points. This gives the required image correspondences for the whole block. Then the matched image correspondences can be reduced before the bundle adjustment, as their number (more than 100,000 image points in some cases) can produce large computational cost for standard PCs. The reduction procedure divides each image in regular cells storing only the point with the best multiplicity. Furthermore, as the extracted features are generally randomly-distributed in the images without a uniform distribution, the method improves also the geometric distribution of the correspondences with a regularization of the extracted and reduced tie points. This strategy leads to a significant reduction (more than 10 times) of the correspondences without loosing precision in the bundle adjustment, and a reduction in the computational time of the exterior orientation parameters (Barazzetti *et al.*, 2010).

3. EXPERIMENTS

The ATiPE procedure was tested on several blocks (Fig. 2) taken by rotary wing UAVs, either with autonomous guidance capability and without (R.C. helicopters).

Block 1 regards some photogrammetric blocks acquired by means of a R.C. helicopter around the roof of the Milan gothic Cathedral (Scaioni *et al.*, 2009). The complexity of the block structure, composed of several tens of unordered convergent images of architectural structures, and with no approximation for the EO parameters, represents a hard test for the capabilities of the developed automated orientation procedure.

A second block (2) was acquired with a mini UAV-system (Copter from Surveycopter) over the archaeological Maya area of Copan, Honduras. The block features a structure similar to an aerial survey, but with much larger scales and strong perspective differences among the images. Different problems were faced here, from the untextured ground area to the occlusions given by the dense vegetation.

A third block (3) was flown over an archaeological site in Siena (Italy), using a 4-rotor Microdrone UAV.

Block 1 – Milan	Block 2 – Copan	Block 3 – Pava	Block 4 – Veio
Nikon D200 20 mm lens – 80 images 3872×2592 pixels pixel size 6.1 μm	Nikon D2Xs 35 mm lens – 70 images 4288×2848 pixels pixel size 5.5 μm	Pentax Optio A40 8 mm lens – 32 images 4000×3000 pixels pixel size 1.9 μm	Pentax Optio A40 8 mm lens – 87 images 4000×3000 pixels pixel size 1.9 μm
Manual flight	Autonomous flight	Manual flight	Autonomous flight
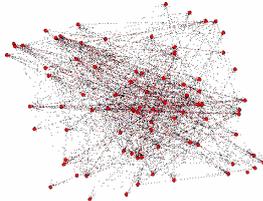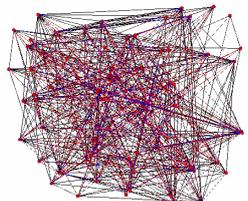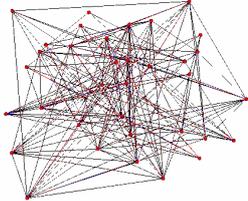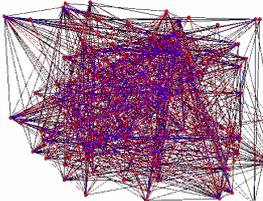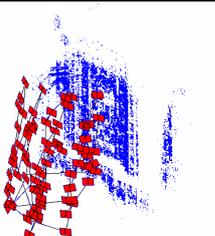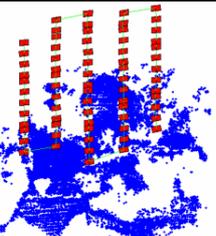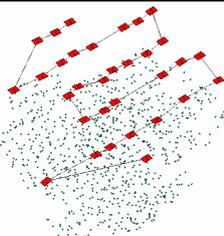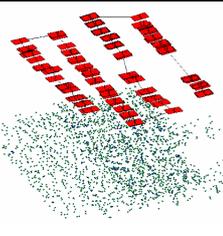
Figure 2. Some features of the image blocks used in the reported experiments: (1) main characteristics of the images; (2) flight mode; (3) example of an image for each block; (4) images of the UAV used; (5) visibility map (red dots are the images and blue lines give their connections); (6) orientation results after the photogrammetric bundle adjustment.

Here the image acquisition phase was manually driven from the ground by an operator with the real-time viewer of the camera frame and GPS/INS values were not available. The geometry of the block resulted quite complex, due to the non-constant overlap. Furthermore, problems raised from the bad/uniform texture of the excavation area.

Lastly, a large image block over the archaeological area of Veio (Italy) with a 4-rotor Microdrone UAV coupled with GPS/INS data was considered (block 4). This project was also used to check the accuracy of the method by using a comparison with some ground check points.

All the presented examples witness the usefulness of the developed automated orientation procedure to support UAV data processing, to reduce the processing time in all cases where the direct orientation is not enough accurate or in case where GPS/INS cannot be used at all.

3.1 Relative orientation with ATiPE

The automated matching phase of each of the blocks with ATiPE took some hours (the smaller block has 32 images with 12 Mpx). In all these experiments the *visibility map* was estimated to reduce the number of image pair combinations. The advantage of this procedure is remarkable, because for each block it leaded to a significant reduction of the computational time (the time is directly proportional to the number of combinations), as can be seen in Table 1.

Block	Global image combinations	Estimated combinations
Block 1	3,160	1,179
Block 2	2,415	507
Block 3	496	170
Block 4	3,741	897

Table 1. The global number of image pairs and those estimated through the visibility map.

The matching phase for each pair is relatively simple for blocks 2, 3 and 4. The geometry of the network is quite similar to standard aerial photogrammetric applications, where there is only a translation and a rotation around the optical axis of the camera. Block 1 is made up of convergent images with strong perspective deformations. However, the SIFT operator was capable of extracting an impressive number of tie points even under extreme viewing conditions (relative rotations up to 30°).

The global number of matched points for all the reported projects is shown in Table 2. The reported numbers of image correspondences are quite huge and in some situations this is a limit for the computation in the Least Squares solution. Furthermore the distribution of the extracted correspondences is often not optimal, with some areas full of points and other without features. Therefore, a reduction and regularization of the extracted tie points becomes necessary to operate with standard PCs and to achieve better adjustment results. Figure 3 shows the results of the developed strategy for an image pair of block 1. The initial distribution of tie points is not good as the points are distributed almost in the whole image, but there is a great redundancy in most of the areas. The reduction was carried out with rectangular cells (the user has to set their size) to remove several useless points. The computed RMS of the bundle adjustment shows how this method does not degrade the precision of the final solution (see Table 2). From a computational point of view, the system with reduced image

points was solved in a few seconds, while in the case of the original dataset the elaboration took several minutes.

Another interesting aspect is related to some mismatches still present after the automated identification of the image correspondences. This is due to the check of the epipolar geometry of an image pairs only: mismatched points lying on the epipolar line cannot be removed with a robust estimation of the *fundamental matrix*. However, a robust bundle solution, where tie points with high residuals are removed by *data-snooping*, can detect and remove these outliers. The effectiveness of the method is improved by the high local reliability of manifold rays that can be viewed on different images. For this reason, the use of a robust bundle adjustment becomes necessary for these applications.

It is also evident how the reduction of tie points simplifies also the computation of the bundle solution. Indeed, only tie points with the largest multiplicity are held.

	Block 1	Block 2	Block 3	Block 4
	without reduction			
image points	120,765	57,477	52,878	331,892
3D points	21,248	27,548	4,040	-
RMS (px)	0.67	0.62	0.79	-
3D outliers	1,207	122	3,513	-
	with reduction			
image points	10,823	6,498	3,513	18,320
3D points	1,207	117	822	2,353
RMS (px)	0.66	0.57	0.8	1.06
3D outliers	3	23	2	34

Table 2. Bundle adjustment statistics with and without point reduction procedure. Block 4 could not be oriented without reduction due to the too much large number of correspondences.
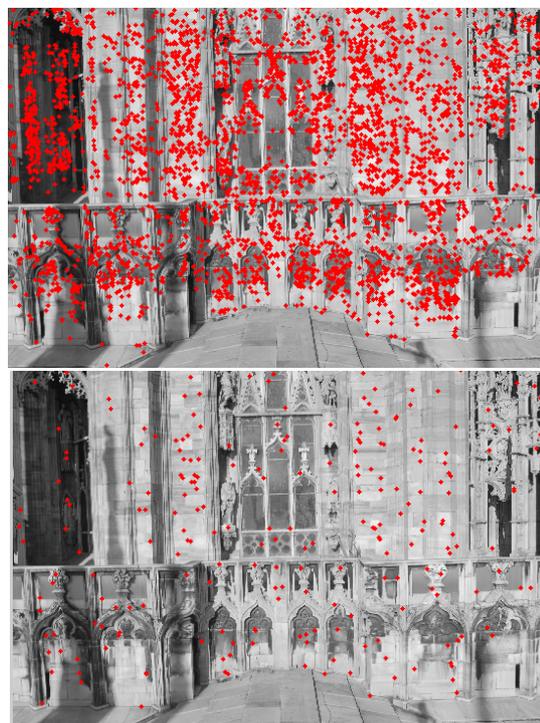
Figure 3. An image pair of block 1 before (top) and after (bottom) the tie point reduction.
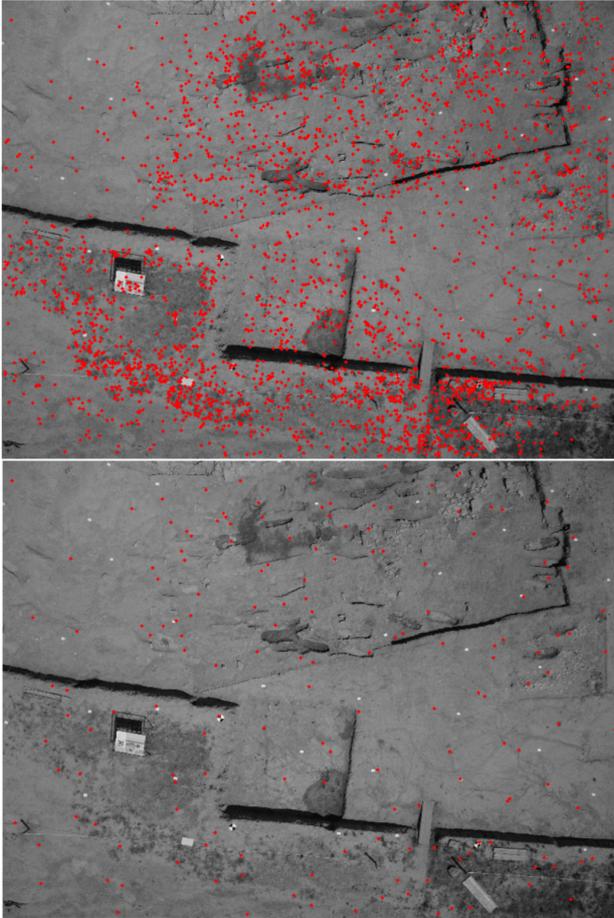
Figure 4. An image pair of block 3 before (top) and after (bottom) the tie point reduction and regularization.

3.2 Accuracy of the method

The statistics of the photogrammetric bundle solution show how this method provides good results in the case of images with a normal configuration, but also when convergent views are employed. However, photogrammetric projects need to be checked and compared with some data acquired with other sensors and with an accuracy better than the one achievable from the image processing.

An accuracy analysis of the developed procedure was carried out for block 4. Some targets, randomly distributed in the scene and measured with a total station, were used as check points. In particular, 3 targets were used as ground control points (GCPs) to estimate a similarity transformation between the photogrammetric project *datum* and the total station *datum*. The remaining targets were employed as check points to estimate the RMSE values. The images of block 4 have an average scale number 1:4,400, with a mean flying height of 35 m (i.e. 0.8 cm footprint). As the block has the typical configuration of aerial projects, an estimation of the theoretical precision along the vertical line can be carried out with the well-known relation (Kraus, 2007):

$$\sigma_z = \pm \frac{Z^2}{cb} \sqrt{2} \sigma_{imm} \qquad (1)$$

where Z is the camera-ground distance, c is the focal length, b is the baseline between two consecutive images and σ_{imm} is the standard deviation of image coordinate measurement. This formula gives out a precision along z equal to ±8 cm.

The computed RMSE on 5 targets of block 4 were:

$$RMSE_x = 4 \text{ cm}; RMSE_y = 3 \text{ cm}; RMSE_z = 7 \text{ cm}$$

that are consistent with the expected theoretical precision.

The analysis of the covariance matrix of the bundle solution gave the following average precisions of the object points (sigma-nought is 1.6 μm, i.e. in the order of the pixel size – see Fig. 2):

$$\sigma_x = \pm 0.6 \text{ cm} ; \sigma_y = \pm 0.6 \text{ cm} ; \sigma_z = \pm 2.3 \text{ cm}$$

that are better than the computed RMSE values. These discrepancies are surely due to the manual measurement of the target coordinates in the images: the centre of each target was difficult to measure (not clearly visible) and some images exhibit a strong blurring.

These results can be improved by using a Least Squares Matching refining (LSM, Gruen, 1985) for all targets but also for natural points. This algorithm is already implemented in ATiPE but, at the moment, it cannot cope with a strong rotation of the camera along its optical axis (e.g. correspondences on different strips). However, in Remondino and Ressl (2006) it was demonstrated how a LSM implementation coupled with the SIFT (or SURF) descriptors can overcome this problem and will be taken into account in further improvements.

4. CONCLUSIONS AND FUTURE WORK

The paper presented a procedure (ATiPE) for the automatic extraction of image correspondences from UAV blocks for their successive orientation with a bundle adjustment solution. Indeed, in most UAV photogrammetric applications the availability of GPS/INS data is not enough to achieve the required precision, exception made when low-quality or small scale products are needed. Consequently, the exterior orientation parameters of the images are computed with an image triangulation, and the automated extraction of the required tie points is necessary to reduce the computing time.

As standard AAT procedures often fail on some UAV blocks because the images feature problems typical of close-range applications (convergent images, large scale, strong perspective deformations, rapid lighting changes, etc.), ATiPE was specifically developed to automatically extract a redundant set of tie points from those blocks.

All the presented examples witness the usefulness of the developed automated procedure to support UAV data processing, to reduce the processing time in all the cases where the direct orientation is not accurate or in the case in which GPS/INS cannot be used at all. Obviously, the developed tie point extraction procedure can be also used when the GPS/INS system is fully operational and only a refinement of the orientation is required. ATiPE, which can be coupled with any existing bundle adjustment solution, differentiates from other methods for its speed, robustness, ability of achieve good distribution of the image correspondences, fully automation and for the high accuracy results obtained at the end of the bundle adjustment. Of course, in the case where metric and geo-referenced results are required, the identification of GCPs in the images still needs to be done manually.

As ATiPE does not require any user interaction, ideally, it would also allow for online triangulation of UAV images.

REFERENCES

References from books

Eisenbeiss, H., 2009. *UAV photogrammetry*. Diss. ETH No. 18515, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland, Mitteilungen Nr.105, p. 235

Hartley R., Zisserman A., 2004. *Multiple View Geometry in computer vision*. Cambridge University Press.

Hofman-Wellenhof, B., Lichtenegger, H. and Wasle, E., 2008. *GNSS – Global Navigation Satellite Systems*. Springer, Wien, New York, 516 pp.

Kraus, K., 2007. *Photogrammetry: Geometry from Image and Laser Scans*. Walter de Gruyter, 459 pp..

Luhmann, T., Robson, S., Kyle, S. and Harley, I., 2006. *Close Range Photogrammetry*. Whittles Publishing, Caithness, Scotland, UK, 510 pp..

References from journals

Arya, S., Mount, D.M., Netenyahu, N.S., Silverman, R. and Wu, A.Y., 1988. An optimal algorithm for approximate nearest neighbour searching fixed dimensions. *Journal of the ACM*, 45(6): 891-923.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. SURF: Speeded up Robust Features. *CVIU*, 110(3), pp. 346-359.

Büyüksalih, G. and Li, Z., 2005. Practical experiences with Automatic Aerial Triangulation using different software packages. *Photogrammetric Record*, 18(102), pp. 131-155.

Gruen, A., 1985: Adaptive least square correlation: a powerful image matching technique. *South African Journal of PRS and Cartography*, 14(3), pp. 175-187.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2), pp. 91-110.

Snavely, S.M., Seitz, R., Szelinski, R., 2008. Modelling the world from internet photo collections. *IJCV*, 80(2), pp. 189-210.

Torr, P., 2002. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *IJCV*, 50(1), pp. 35-61.

Zhou, G., 2009. Near Real-Time Orthorectification and Mosaic of Small UAV Video Flow for Time-Critical Event Response. *IEEE Trans. Geoscience and Remote Sensing*, 47(3), pp. 739-747.

References from proceedings

Barazzetti, L., Remondino, F., and Scaioni, M., 2009. Combined use of Photogrammetric and Computer Vision techniques for fully automated and accurate 3D modeling of terrestrial objects. In: Proc. of Int. Conf. "Videometrics, Range Imaging, and Applications X", 2-3 Aug, San Diego (CA, USA), *SPIE*, Vol. 7447, 12 pp.

Barazzetti, L., Remondino, F., and Scaioni, M., 2010. Automation in 3D reconstruction: results on different kinds of close-range blocks. *IAPRS&SIS*, Vol. 38, Part 5, Newcastle u. Tyne, UK (in press).

Colomina, I., Blázquez, M., Molina, P., Parés, M.E. and Wis, M., 2008. Towards a new paradigm for high-resolution low-cost photogrammetry and remote sensing. In: *IAPRS&SIS*, Vol. 37, Part B1, Beijing, China, pp. 1201-1206.

Eisenbeiss, H., 2008. The autonomous mini helicopter: a powerful platform for mobile mapping. In: *IAPRS&SIS*, Vol. 37, Part B1, Beijing, China, pp. 977-983.

Eugster, H. and Nebiker, S., 2008. UAV-based augmented monitoring – real-time georeferencing and integration of video imagery with virtual globes. In: *IAPRS&SIS*, Vol. 37(B1), Beijing, China, pp. 1229-1235.

Grenzdörffer, G.J., Engel, A. and Teichert, B., 2008. The photogrammetric potential of low-cost UAVs in forestry and agriculture. In: *IAPRS&SIS*, Vol. 37, Part B1, Beijing, China, pp. 1207-1213.

Pueschel, H., Sauerbier, M., Eisenbeiss, H., 2008. A 3D model of Castle Landemberg (CH) from combined photogrammetric processing of terrestrial and UAV-based images. In: *IAPRS&SIS*, Vol. 37, Part B6-b, Beijing, China, pp. 96-98.

Remondino, F., Ressl, C., 2006. Overview and experience in automated markerless image orientation. *IAPRSSIS*, Vol. 36(3), Bonn, Germany, pp. 248-254.

Scaioni M., Barazzetti L., Brumana R., Cuca B., Fassi F., Prandi F., 2009. RC-Heli and Structure & Motion Techniques for the 3-D Reconstruction of a Milan Dome Spire. In: *IAPRSSIS*, Vol. 38, Part 5/W1, Trento, Italy (on CDROM).

Wang, J., Garratt, M., Lambert, A., Wang, J.J., Han, S. and Sinclair, D., 2008. Integration of GPS/INS/vision sensors to navigate unmanned aerial vehicles. In: *IAPRS&SIS*, Vol. 37, Part B1-b, Beijing, China, pp. 963-969.