EXPLOITING MULTI-CAMERA CONSTRAINTS WITHIN BUNDLE BLOCK ADJUSTMENT: AN EXPERIMENTAL COMPARISON

E. Maset[1], E. Rupnik[2], M. Pierrot-Deseilligny[2], F. Remondino[3], A. Fusiello[4,*]

[1] Dept. of Agricultural, Food, Environmental and Animal Sciences, University of Udine, Udine, Italy - eleonora.maset@uniud.it
[2] LASTIG, Univ. Gustave Eiffel, ENSG, IGN, F-94160 Saint-Mande, France - ewelina.rupnik@ign.fr, marc.pierrot-deseilligny@ensg.eu
[3] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy - remondino@fbk.eu
[4] Polytechnic Department of Engineering and Architecture, University of Udine, Udine, Italy - andrea.fusiello@uniud.it

Commission II, WG II/1

KEY WORDS: Bundle block adjustment, multi-head camera systems, oblique images, constrained cameras, orientation

ABSTRACT:

The growing deployment of multi-head camera systems encouraged the emergence of specific processing algorithms, able to face the challenges posed by slanted view geometry. Such multi-camera systems are rigidly tied by their manufacturers hence the exploitation of this internal constraint should be further exploited. Several approaches have been proposed to deal with orientation constraints, with the aim of reducing the number of unknowns, computational time and possibly improve the accuracy. In this paper we compare the results provided by publicly available implementations in order to further investigate the advantages of enforcing relative orientation constraints for aerial and terrestrial triangulation of multi-head camera systems. Data from a Leica CityMapper and a Stereopolis-Ladybug are considered, reporting how constrained solution can improve accuracy with respect to traditional (unconstrained) bundle block adjustment solutions.

1. INTRODUCTION

In the last years, nearly all existing companies in the geospatial industry have embraced multi-camera oblique imaging technology thereby expanding the potential of the area-wide mapping market. The use of multiple cameras is also commonly employed for SLAM (*Simultaneous Localisation and Mapping*) in stereo configurations or as omni-directional cameras (Kaess and Dellaert, 2006). Their benefit is the instantaneous provision of the scaled 3D geometry, as well as the extended field-of-view, which simplify the localisation task. In the aerial context, multi-head camera systems provide the advantages of slanted view geometry, which allows for the 3D reconstruction of building facades and other vertical objects (Haala and Rothermel, 2015). However, this poses new challenges, which include dealing with image scale variability, multiple occlusions, and greater disparity in search space.

Regarding the image orientation problem, several works in the literature suggest that relative orientation constraints among the cameras should be considered (Wiedemann and Moré, 2012, Rupnik et al., 2015), in order to reduce the number of unknowns and possibly to improve the accuracy. In this regard, two main approaches have been proposed to deal with orientation constraints:

- The first one, frequently applied at commercial level for terrestrial and aerial multi-head camera rigs, consists in recovering the relative orientations between the cameras during a calibration procedure (Esquivel et al., 2007, Dai et al., 2009, Schneider and Förstner, 2013). After this initial step, a bundle block adjustment (BBA) will optimize only the exterior orientation parameters of a reference camera, the others being rigidly connected to the latter. Alternatively, the calibrated rig can be considered as a single,

non-perspective camera (i.e., a camera in which the bundle of rays do not intersect in a unique point), and the exterior orientation of the multi-head system is cast as a *Non-perspective-n-Point* problem (Fusiello et al., 2015).

- The second possible approach is to bypass the preliminary calibration and compute the relative orientations among the cameras directly from the data. This is done, with different nuances, in COLMAP (Schönberger and Frahm, 2016), MicMac (Pierrot-Deseilligny et al., 2014) and CRO-BBA (Maset et al., 2020).

The aim of this work is to compare the results provided by the publicly available implementations of the above mentioned methods in order to further investigate the advantages of enforcing relative orientation constraints for aerial and terrestrial triangulation of multi-head camera systems.

2. METHODS

Let us consider a rigid multi-head system composed of k synchronized cameras, where one camera is taken as the *reference*, and the remaining $k - 1$ *underling* ones have a fixed but unknown relative orientation with respect to the reference one. Interior orientation parameters are known and fixed.

All the three considered methods starts from an initialization obtained from Structure-from-Motion (SfM) and infer the relative poses from the data, as opposed to other methods that take the relative poses from calibration.

The `RigBundleAdjuster` (RBA in the following) of **COLMAP** (Schönberger and Frahm, 2016) computes the average relative orientations between rigged cameras from the initial SfM and then considers them known in the final BBA, as in a calibrated case (Heng et al., 2015). Please note

* Corresponding author

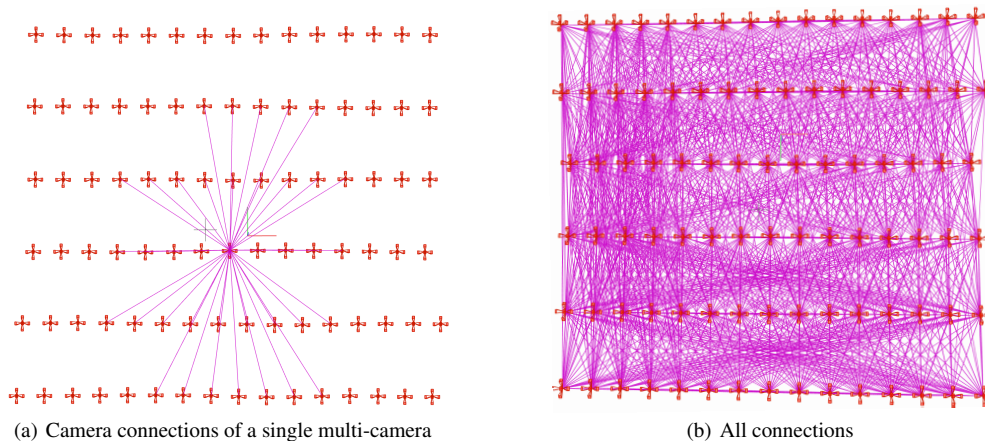(a) Camera connections of a single multi-camera   (b) All connections

Figure 1. Leica CityMapper dataset. The two plots depict the overlap pattern of the images. In red are the camera poses, and in magenta are the image connections.

that they are introduced as (hard) constraints forced at each iteration, so the number of unknowns remain the same as in the unconstrained version of BBA.

The **MicMac** (Pierrot-Deseilligny et al., 2014) approach, called `Rigid Block Compensation` – RBC in the following – computes the initial relative orientations by averaging from the initial SfM. Then, it parametrises the multi-head cameras with their independent exterior orientation parameters, and applies the relative orientations as soft constraints, thereby relaxing the rigidity of the rig (as opposed to the other two methods, where the rigidity is strictly enforced)[1]. It provides a weighting scheme that allows the relative orientation to remain constant, evolve from its initial value up to a certain value or evolve over time. Unlike CRO-BBA (Maset et al., 2020), and similar to RBB (Schönberger and Frahm, 2016), the RBC's parametrisation does not reduce the number of BBA unknowns. For more detailed information on the implementation and the use of the RBC, refer to the software's manual (Pierrot-Deseilligny et al., 2014).

The **CRO-BBA** algorithm (Maset et al., 2020) introduces the fixed relative orientations among the rigged cameras as unknowns in the BBA that is customarily run as the last stage of the SfM pipeline. It expresses the exterior orientation parameters of the $k - 1$ underling cameras as a function of the parameters of the reference camera and of the relative orientations (fixed but unknowns). In this way, the rigidity of the multi-camera system is enforced while computing the unknown relative orientations among the rigged cameras. The method implements the exact formulation of the Jacobian matrix, that collects the partial derivatives of the collinearity equations rewritten to account for relative constraints.

The Jacobian of the classical BBA (with known interior orientation) is composed of blocks of two types, that contains the derivatives of the residuals wrt exterior orientation parameters and 3D point coordinates, respectively. In the CRO-BBA these blocks correspond to the reference camera, whereas for the underling cameras three new block types are introduced: i) the derivative of the residual wrt the orientation of the reference camera; ii) the derivative of the residual wrt the relative orientation; iii) the derivative of the residual wrt the 3D point coordinates.

CRO-BBA retrieves all the orientations via a free-network adjustment with the Levenberg-Marquardt optimization strategy (which implicitly removes the datum defect).

3. MATERIALS

In this study we considered two datasets as presented in Tab. 1:

Camera system	Sensor [pix]	Focal [mm]	# images
Leica CityMapper$_{nadir}$	10336 x 7788	83	92
Leica CityMapper$_{oblique}$	10336 x 7788	156	368
Stereopolis$_{Ladybug}$	2528 x 2080	4	3245

Table 1. Characteristics of the cameras used in the experiments.

Aerial dataset (Leica CityMapper). This dataset (Toschi et al., 2018) is composed of 460 images (92 nadir images and 368 oblique images) collected by the Leica CityMapper hybrid sensor over the city of Heilbronn, Germany (data courtesy by Leica Geosystems). Leica CityMapper combines a Hyperion LiDAR unit (1064 nm wavelength, theoretical ranging accuracy <2 cm) and a multi-camera system, featuring one nadir-looking camera head (10,336 × 7,788 pixels, 83 mm focal length) and four $45°$-tilted camera heads. The flight plan was designed using an average nadir GSD (ground sample distance) of 12 cm, and along-across overlaps of 80% and 60%, respectively. The selected subset covers an area of ca. 3.5 km × 3.5 km. Fig. 1 shows the connections among overlapping images.

Terrestrial dataset (Stereopolis-Ladybug). The images were acquired with the *Stereopolis* mobile mapping platform (Paparoditis et al., 2012). The camera system is composed of five Ladybug cameras equipped with fisheye lenses, rigidly installed on top of a car. The cameras were arranged symmetrically around the car's vertical axis so as to give a $360°$ scene coverage. All five cameras were prior calibrated on a 3D test calibration field. The data acquisition protocol of the experiment is presented in Fig. 2. To be able to asses the trajectory's drift caused by the unmodelled error accumulation, a 2 km long trajectory was simulated by driving the car around the same block of buildings in eight rounds.

COLMAP[2], MicMac[3] and CRO-BBA[4] implementations are publicly available on the web.

---

[1] MicMac is also capable of including pre-calibrated relative orientations in the processing, but this is not relevant to this comparison

[2] https://github.com/COLMAP/
[3] https://github.com/micmacIGN
[4] http://www.dpia.uniud.it/fusiello/demo/oba/

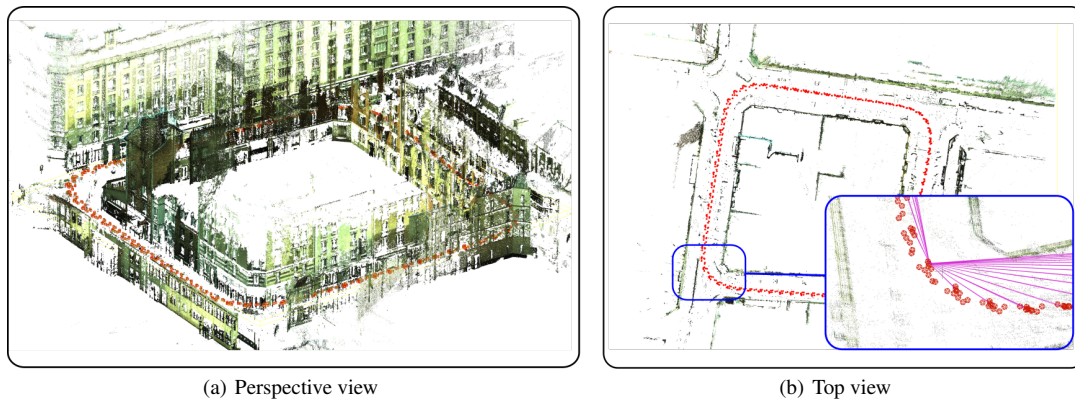(a) Perspective view   (b) Top view

Figure 2. Stereopolis-Ladybug trajectory. To simulate a long-distance trajectory, the car drove in circles around the block of buildings. In red are the camera poses, and in magenta are the image connections.

4. EXPERIMENTS

4.1 Evaluation protocol

Leica CityMapper. The same set of 8000 tie-points employed in (Maset et al., 2020) was used in all the methods, and the results were evaluated on 49 Ground Control Points (GCPs), whose 3D coordinates were measured by means of RTK Global Navigation Satellite System (GNSS) with 5 cm accuracy. More in detail, the commercial software 3DF Zephyr was used to detect keypoints and compute the corresponding descriptors following a SIFT-like approach. A robust matching was then performed and the longest 8000 tracks were fed to the evaluated methods. Please note that tie-points were extracted on half-resolution images, so the effective GSD is 24 cm.

For the purpose of the comparisons conducted in this research, bundle adjustment was carried out in an arbitrarily defined datum, and interior parameters were considered known (we adopted the values reported in the calibration certificate of the system). Errors in object-space are eventually computed as the Root Mean Square (RMS) of the residual distances between corresponding 3D points after least-squares (Procrustes) alignment of object points to GCPs. Since all the methods produce free-network solutions and the alignment transformation is a similitude (a.k.a. Helmert transformation), any non-rigid deformation of the model is revealed (in object-space) by the alignment residuals.

Stereopolis-Ladybug. The reference trajectory was obtained by photogrammetric processing with GCPs (6 points measured by a total station with $\sigma < 1$ cm), and by exploiting the closed loops (i.e., tie-points across the rounds). The 6 GCPs practically become 48 GCPs because every round is considered independent. The accuracy of the reference trajectory calculated as a mean distance between the GCPs positions predicted by photogrammetry and their ground truth positions is equal to 5.5 cm.

To test the influence of the constrained BBA only tie-points between consecutive images were extracted (cf. image connections in Fig. 2). The experiments were conducted in an arbitrary reference frame, and were compared to the reference trajectory after applying a Helmert transformation. The first 20 acquisitions (i.e., 100 images) were used to computed the optimal transformation. Figure 6 illustrates the distances between the computed camera poses and the reference trajectory.

The same set of tie-points was used in both, MicMac and COLMAP. The points were extracted in COLMAP using the

`sequential_matcher` program. Each image was matched with up to 24 images before and after, amounting to extracting 10,409,315 tie-points for the entirety of the dataset. To filter out the false matches found within the "static" parts of the image (e.g., the rig installation visible in each image), as well as the matches over the sky, per-image masks were applied.

Note that the Ladybug was calibrated using MicMac software, whose fisheye camera model is different from the one proposed in COLMAP. We therefore allowed COLMAP to refine the initial internal calibrations.

CRO-BBA is not included in this comparison as it cannot handle large numbers of images × tie-points, being implemented in Matlab.

4.2 Results

Leica CityMapper. Results are illustrated in Fig.s 3 and 4, which report respectively planimetric (XY) and altimetric (Z) components of object-space errors for GCPs, for the three considered methods. Figure 5 shows the boxplots of the statistics of 3D object-space errors. The central mark of each box represents the median value, while the bottom and top edges correspond to the 25th and 75th percentiles, respectively. Lines extending from the box indicate the most extreme data not considered as outliers.

For all the evaluated methods, the constrained solution provides more accurate results than the corresponding classical BBA, as one would expect. The RMS error computed on the 49 GCPs is very similar for the three methods: 0.15 m for COLMAP, 0.16 m for MicMac and 0.14 m for CRO-BBA, whereas in the unconstrained case (BBA) the RMS error increases to 0.25 m, 0.22 m and 0.20 m, respectively. Enforcing constraints impacts both the planimetric and altimetric precision of the 3D measurements.

Stereopolis-Ladybug. The results are presented in Fig. 6. The introduced constraints effectively help to decrease the drift errors. In MicMac, the RMS errors across the ≈ 2 km long trajectory are 0.67 m with the enforced rigidity, and 1.41 m otherwise. Similarly, the RMS for COLMAP is 0.64 m with or 2.90 m without the constraints. The planimetric accuracy is more impacted than the altimetry when applying the rigidity constraints. Note that the given accuracies correspond to purely photogrammetric performance. In practice, much better accuracies can be obtained by coupling photogrammetry with

(a) COLMAP

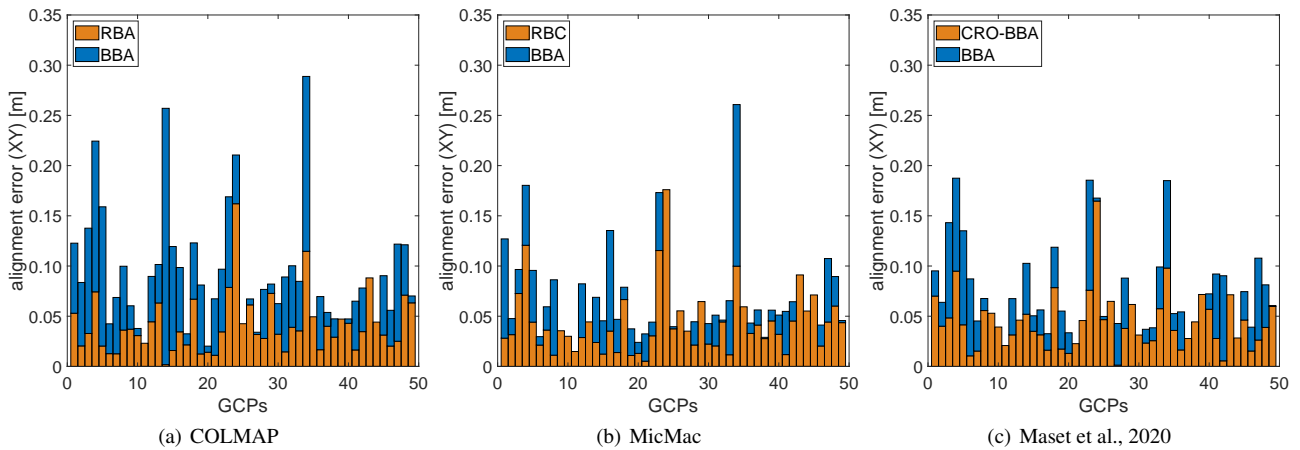(b) MicMac

(c) Maset et al., 2020

Figure 3. Results of the comparison experiments for the Leica CityMapper dataset between the unconstrained (blue bars) and constrained (orange bars) bundle adjustments. Planimetric component of object-space errors for GCPs.
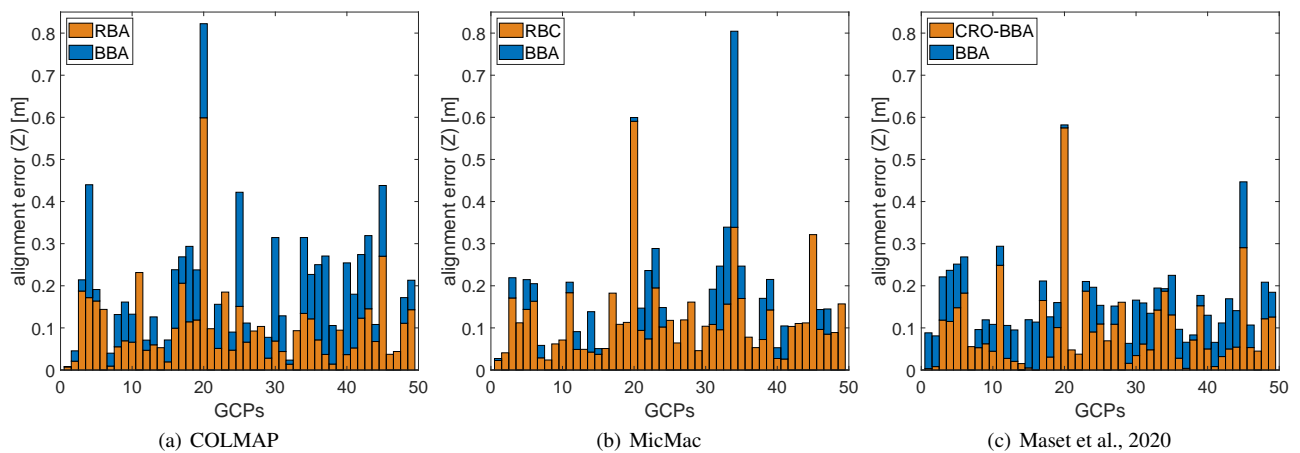

(a) COLMAP

(b) MicMac

(c) Maset et al., 2020

Figure 4. Results of the comparison experiments for the Leica CityMapper dataset between the unconstrained (blue bars) and constrained (orange bars) bundle adjustments. Altimetric component of object-space errors for GCPs,
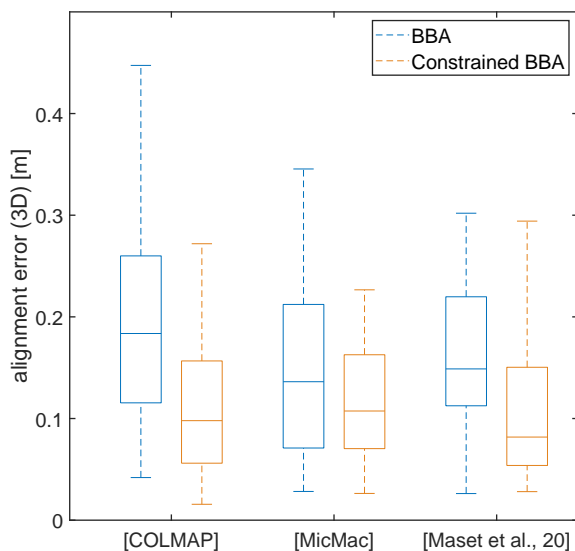

Figure 5. Leica CityMapper: boxplot of 3D object-space errors provided by the three considered methods.

either GCPs or GNSS/IMU. What the presented results reveal, however, is that the rigid constraints slow down the drift, and as a consequence, fewer GCPs or GNSS/IMU points would be needed to eliminate it. This is crucial because GCP collection is labour intensive, and GNSS in urban corridors is often unreliable.

The characteristic ripple along the trajectory in Fig. 6 is an artefact caused by the repetitive geometry of the trajectory.

5. DISCUSSION

The results confirm the effectiveness of enforcing relative orientation constraints in the BBA of multi-head camera systems, showing comparable performance for the tested algorithms.

Soft versus hard constraints. The two tested acquisition systems are high-end systems with good mechanical stability and precise camera synchronisation. As a result, the soft constraint available in MicMac could not be evaluated. We found that optimal results were obtained by not allowing the rig to evolve over time, and when fixing the relative orientation weights to conservative values. In the experiments, the following weights on the rotational component σ_R and the translation σ_T of the camera rig were adopted: Leica Citymapper – $\sigma_T = 1e^{-3}$ [m], $\sigma_R = 1e^{-5}$ [rad]; Stereopolis-Ladybug – $\sigma_T = 1e^{-3}$ [m], $\sigma_R = 1e^{-4}$ [rad].
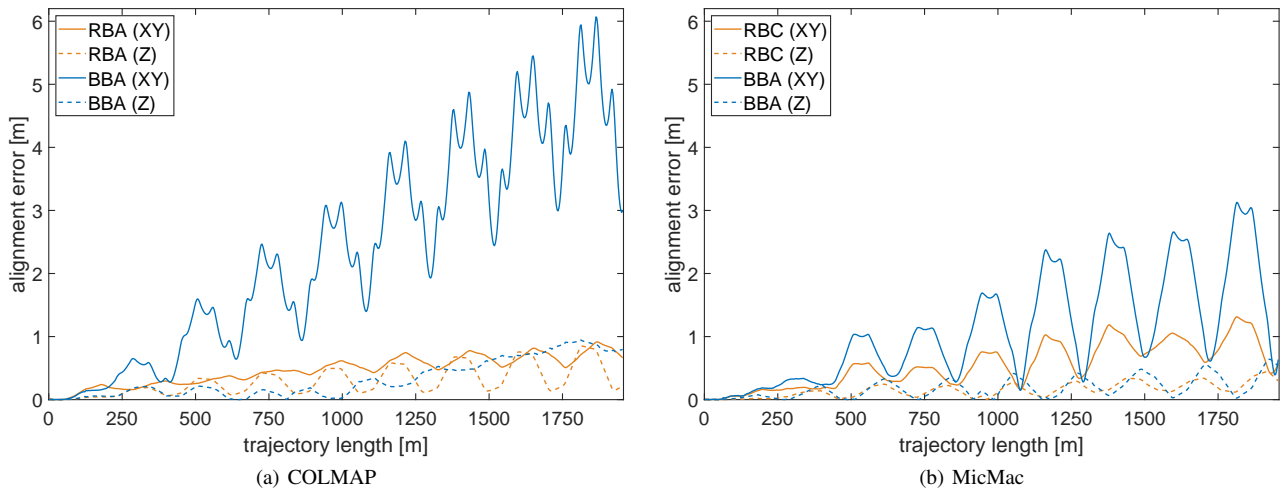
(a) COLMAP

(b) MicMac

Figure 6. Results of the experiments for the Ladybug dataset. Distances between estimated camera poses and ground truth trajectory, produced (a) by COLMAP and (b) by MicMac. Solid lines show the planimetric residuals, while dashed lines show the altimetric component.
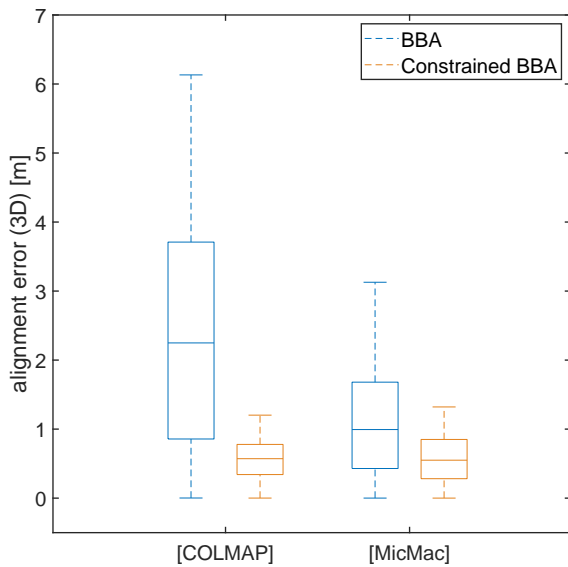

Figure 7. Stereopolis-Ladybug: boxplot of object-space errors provided by the two considered methods.

Correlations between parameters. We also show empirically that the correlations between external orientations decreased significantly when the rigidity was imposed. Figure 8 presents the correlations coefficients between a pair of images in two scenarios: the classical BBA and the constrained BBA. The coefficients were calculated from the co-variance matrices Σ as $\frac{\Sigma_{i,j}}{\sqrt{\Sigma_{i,i} \cdot \Sigma_{j,j}}}$.

6. CONCLUSIONS

The paper presented an investigation on multi-head camera systems to compare the performances of publicly available image orientation processes. We evaluated three different implementations (COLMAP, MicMac, CRO-BBA) of constrained bundle block adjustment which impose a degree of rigidity over multi-head camera systems. The tested implementations differ in two aspects:

(a) Leica CityMapper
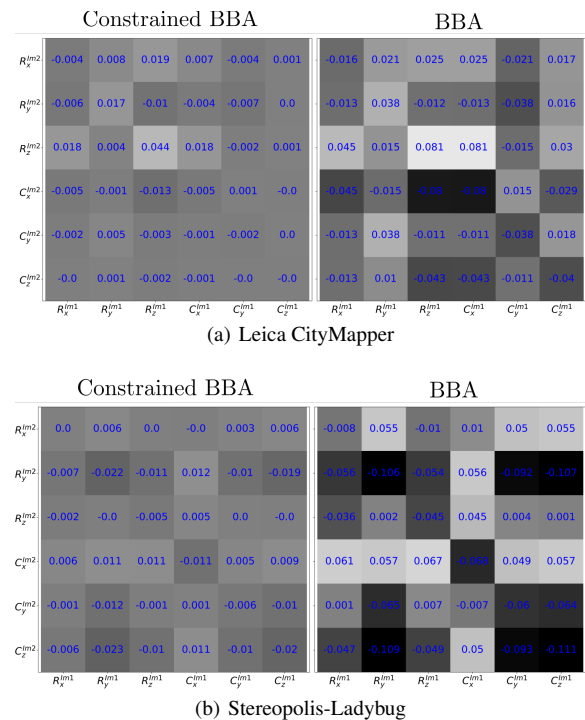

(b) Stereopolis-Ladybug

Figure 8. Correlation matrices between two sample images in (a) Leica CityMapper and (b) Stereopolis-Ladybug datasets. In each 6x6 matrix, the first three elements denoted as R_x, R_y, R_z correspond to the rotations, whereas the latter C_x, C_y, C_z correspond to the perspective center. The left column presents the result of the constrained BBA, and the right column is the BBA without constraints.

- parametrization of the cameras, i.e., with independent exterior orientations for each camera (COLMAP, MicMac) or with a reference camera and relative orientations for the underling ones (CRO-BBA);

- the degree of rigidity, i.e., imposed as a constant (hard constraint) or as a parameter (soft constraint).

We conducted experiments using aerial as well as terrestrial

datasets and found that all three implementations performed equally well by significantly reducing the residuals in object space. The accuracy gain due to the adopted constraints depends on the geometry of the acquisition (i.e., improvement on all three coordinates in the aerial case, as opposed to an improvement only in planimetry for the terrestrial case). We could not asses the advantage of the approach proposing the soft constraints over the hard constraints because the tested camera systems turned out to be rigid. Results are so encouraging further utilization of such constrained bundle adjustments in order to speed up processing time, reduce the number of unknowns and the risk of adjustment's failure.

ACKNOWLEDGMENTS

REFERENCES

Dai, Y., Trumpf, J., Li, H., Barnes, N., Hartley, R., 2009. Rotation averaging with application to camera-rig calibration. *Asian Conference on Computer Vision*, Springer, 335–346.

Esquivel, S., Woelk, F., Koch, R., 2007. Calibration of a multi-camera rig from non-overlapping views. *Joint Pattern Recognition Symposium*, Springer, 82–91.

Fusiello, A., Crosilla, F., Malapelle, F., 2015. Procrustean point-line registration and the NPnP problem. *Proceedings of the International Conference on 3D Vision (3DV)*, IEEE, 250–255.

Haala, N., Rothermel, M., 2015. Image-based 3d data capture in urban scenarios. *Proc. Photogrammetric Week 2015*, 119–130.

Heng, L., Furgale, P., Pollefeys, M., 2015. Leveraging Image-based Localization for Infrastructure-based Calibration of a Multi-camera Rig. *Journal of Field Robotics*, 32(5), 775 - 802.

Kaess, M., Dellaert, F., 2006. Visual slam with a multi-camera rig. Technical report, Georgia Institute of Technology.

Maset, E., Magri, L., Toschi, I., Fusiello, A., 2020. Bundle Block Adjustment with Constrained Relative Orientations. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020, 49–55.

Paparoditis, N., Papelard, J.-P., Cannelle, B., Devaux, A., Soheilian, B., David, N., Houzay, E., 2012. Stereopolis II: A multi-purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology. *Revue française de photogrammétrie et de télédétection*, 200(1), 69–79.

Pierrot-Deseilligny, M., Jouin, D., Belvaux, J., Maillet, G., Girod, L., Rupnik, E., Muller, J., Daakir, M., Choqueux, G., Deveau, M., 2014. Micmac, Apero, Pastis and Other Beverages in a Nutshell. *Institut Géographique National*.

Rupnik, E., Nex, F., Toschi, I., Remondino, F., 2015. Aerial multi-camera systems: Accuracy and block triangulation issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 233 - 246.

Schneider, J., Förstner, W., 2013. Bundle Adjustment and System Calibration with Points at Infinity for Omnidirectional Camera Systems. *Photogrammetrie-Fernerkundung-Geoinformation*, 2013(4), 309–321.

Schönberger, J. L., Frahm, J.-M., 2016. Structure-from-Motion Revisited. *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Toschi, I., Remondino, F., Rothe, R., Klimek, K., 2018. Combining Airborne Oblique Camera and Lidar Sensors: Investigation and New Perspectives. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1, 437–444.

Wiedemann, A., Moré, J., 2012. Orientation strategies for aerial oblique images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39, 185–189.