2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Optimization of Radial Distortion Self-Calibration for Structure from Motion from Uncalibrated UAV Images

Yonglu Li*, Yinghao Cai†, Dayong Wen‡, Yiping Yang§
Institute of Automation, Chinese Academy of Sciences
Beijing, China
Email: *liyonglu2014@ia.ac.cn, †yinghao.cai@ia.ac.cn, ‡dayong.wen@ia.ac.cn, §yiping.yang@ia.ac.cn

***Abstract*—Structure from motion (SfM) and self-calibration from images of unknown radial distortions could fail under some critical configurations and produce distorted reconstruction results. In this paper, we propose an effective approach to optimize the estimation of radial distortion coefficient by taking full advantage of GPS information, which allows for more accurate SfM results. A feedback function is designed as the metric to indicate the magnitude of the distortion error. Heuristic search strategies are applied to search for the optimal distortion coefficient. Extensive experimental results show that our approach can effectively reduce the distorted deformation error and improve the estimation accuracy of the distortion coefficient.**

I. Introduction

Structure from motion (SfM) as an active topic has received much attention in the past few decades [1]–[7]. Recently, Unmanned aerial vehicles (UAVs) are increasingly used for large-scale 3D reconstruction. In UAV applications, cameras usually shoot vertically and move parallel to the ground. The near-parallel imaging conditions and inaccurate self-calibration of unknown radial distortions may produce distorted reconstruction results as shown in Fig. 1. Fig. 1 shows that with accurate or optimal estimation of the radial distortion coefficient, the distortions of the reconstruction could be negligible. With increase/decrease of the distortion coefficient, the results may exhibit central doming deformations. The further away from the optimal estimation, the more severe the doming deformation.

With the development of sensors, geospatial metadata at a fine granular level can be obtained at recording time with built-in Global Positioning System (GPS). GPS information (longitude, latitude and altitude) is widely used for route planning in many UAV applications. We propose a novel method to mitigate the distortion error of SfM from uncalibrated UAV images by taking full advantage of GPS information in this paper. A feedback function is designed as the metric to indicate the magnitude of the distortion error. Heuristic search strategies are then applied to find the optimal distortion coefficient. Finally, partial bundle adjustment with constant distortion coefficient is used after each search step of distortion coefficient to refine the reconstruction and update feedback function iteratively until the metric meets certain requirements. Our method is inspired from the self-optimizing control structure [8] in industrial automation, which the strategy is to find controlled variables and keep them at constant setpoints to achieve near-optimal operation of the system.

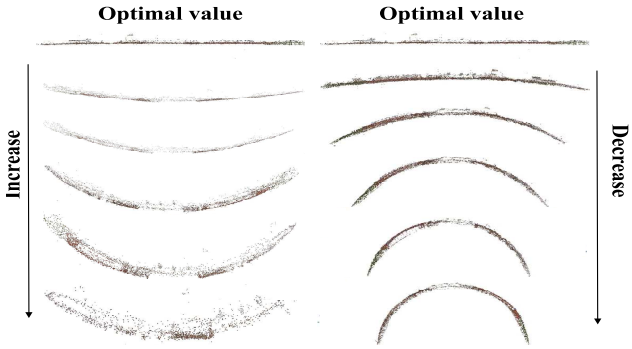


Fig. 1. The near-parallel imaging conditions and inaccurate self-calibration of unknown radial distortions may produce distorted reconstruction. The first row shows that with the optimal estimation of the radial distortion coefficient, the distortion error could be negligible. With increase/decrease of the coefficient, the results may exhibit central doming deformations.

The remaining parts of the paper is organized as follows. An overview of the related work and the ambiguity inherited in radial distortion self-calibration is presented in Section 2. Section 3 introduces our approach of self-calibration of unknown radial distortions, which allows for more accurate SfM results. SfM results with images captured from different flight plans are analyzed and evaluated in Section 4. Ultimately, conclusion and future work are given in Section 5.

II. Related Work

In SfM, images are operated either incrementally or globally. Incremental methods such as Bundler [1] and VSFM [2] start with some seed images and gradually add related images to recover the structure of the scene. Bundle adjustment (BA) [9]–[11] is frequently applied in incremental methods to refine the results. On the contrary, global methods [3]–[7], [12]

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

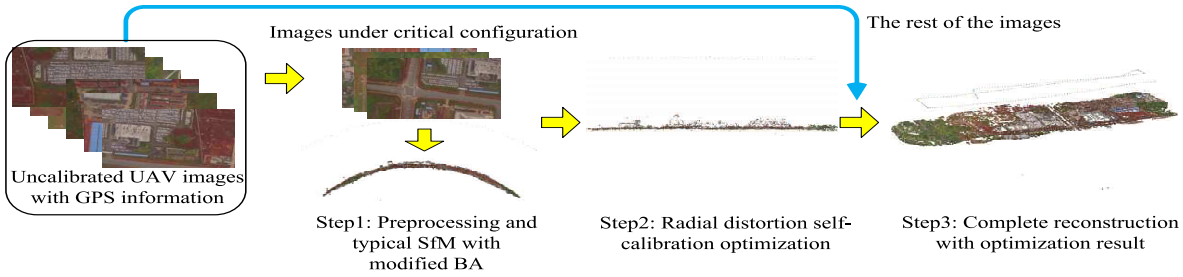


Fig. 2. The flowchart of SfM with unknown radial distortions. The first step is preprocessing and initial reconstruction by typical SfM tools. Heuristic search strategies are applied to search for the optimal value of the radial distortion coefficient in the second step. Finally, reconstruction is rerun on all images using the optimized results from the second step.

simultaneously operate on all the images where BA is applied only once in the final refinement.

The problem of distorted reconstruction caused by inaccurate radial distortion estimation is first raised by Wu [13]. For certain configurations of motion and structure, 3D reconstruction has inherit ambiguities. These configurations are called critical configurations in [13]. A comprehensive study of critical configurations is covered in [13]. It is advised to avoid special capture mode or do pre-calibration in [13]. No effective solution to mitigate the distorted reconstruction is given in [13]. In addition, small camera motions (short baseline) between captures will bring larger accumulated ambiguous errors [13]. In large-scale applications, it is impracticable to calibrate all cameras manually and regularly. To this end, Javenick et al. [14] use many ground control points (GCPs) to rectify the results at the expense of more computing complexity. Wackrow et al. [15] propose to add more oblique captures to mitigate the reconstruction error. GCPs may not easy to obtain in real applications. Moreover, adding oblique images means less efficient in aerial survey since flight plans that arranged in line or block are the most common ways in UAV-based applications.

Many approaches have been proposed to use auxiliary information for SfM [3], [4], [7], [16]. Irschara et al. [16] use GPS locations as the initial camera locations directly. Crandall et al. [3] use the noisy GPS information and vanishing points in the reconstruction process. Guo et al. [7] compute the image overlap rate in image matching using GPS and DEM reference. Recently, Cui et al. [4] propose a framework to exploit noisy sensor data to average camera rotation and initialize the camera center. It is worth mentioning that few attempt has been carried out to mitigate the distorted reconstruction error. In this paper, we propose a new method to estimate the radial distortion coefficient using auxiliary GPS information, which allows for more accurate SfM results. Experimental results on various uncalibrated UAV images show that our approach can accurately estimate the radial distortion and effectively reduce the reconstruction error.

III. Radial Distortion Self-Calibration Optimization

As we mentioned earlier, inaccurate estimation of the radial distortion may produce distorted reconstruction results. In many SfM tools [1], [2], [5], [17], radial distortion is formulated as Brown's distortion model [18]. However, there might be more ambiguities if more radial distortion parameters are used in SfM [13]. For simplicity, we formulate the radial distortion model as:

$$f(r^2) = 1 + k_1 r^2 \tag{1}$$

where r denotes the distortion radius. Only k_1 is used in (1). The flowchart of our approach is shown in Fig. 2. Our approach consists of three steps: preprocessing and initial reconstruction by typical SfM, distortion self-estimation optimization and complete reconstruction using optimized distortion coefficient.

A. *Step* 1*: Preprocessing and SfM with modified BA*

Our approach handles SfM with unknown image radial distortions. We first find images under critical configuration which means that the altitudes of camera in GPS tags are nearly equal, and run initial reconstruction. Since most of the built-in GPS on UAVs are noisy, we do not use GPS data directly as initial camera locations as in [16]. A modified BA cost function consists of reprojection error term and GPS penalty term [4] is used during this process. The rough GPS information can help adjust the camera positions preliminary. The cost function is formulated as:

$$E'(P,E) = \sum_{i=1}^{n}\sum_{j=1}^{m} \lambda_{ij} \|m_{ij} - p(P_i, X_j)\|^2 + \sum_{i=1}^{n} \eta * (C_i - C_i^{GPS})^2 \tag{2}$$

where m_{ij} is the 2D image point location in the i_{th} image. X_j denotes the 3D point. $\lambda_{ij} = 1$ if X_j can be observed in the i_{th} image, otherwise $\lambda_{ij} = 0$; $p(P_i, X_j)$ denotes the projection of the 3D point X_j in the i_{th} image. η is the compensation factor to adjust the influence of GPS penalty term and should

be set relatively small under low GPS accuracy. The three-dimensional camera location of the i_{th} image is denoted as C_i. C_i^{GPS} is the three dimensional location of the i_{th} image in GPS tag. Instead of optimizing on all dataset, we choose to operate on a small set of representative images with ambiguities. Since BA is time-consuming and we will invoke BA frequently in the next step. Further, we use parallel BA approach [11] to accelerate the process.

B. Step 2: Radial distortion coefficient optimization

The main goal of radial distortion coefficient optimization is to find the optimal coefficient to weaken or eliminate the distortion error. A metric is necessary to indicate the magnitude of reconstruction error. Here, the feedback function indicating the reconstruction error at the k_{th} iteration on images selected in step 1 is formulated as:

$$\begin{aligned} F_{metric}^{(k)} = \omega_1 \| Var_C^{(k)} - Var_{C^{GPS}} \| \\ + \omega_2 \| (\sum_i^l max_l\{C_i^{(k)}\} - \sum_i^l min_l\{C_i^{(k)}\}) \\ - (\sum_i^l max_l\{C_i^{GPS}\} - \sum_i^l min_l\{C_i^{GPS}\}) \| \end{aligned} \quad (3)$$

where Var is the variance of camera altitudes distribution, which represent the degree of dispersion of camera locations in the z-coordinate. $Var_C^{(k)}$ and Var_C^{GPS} denote the variance of estimated altitudes at the k_{th} iteration and recorded altitudes in GPS tags, respectively. $\sum_i^l max_l\{a_i\} - \sum_i^l min_l\{a_i\}$ denotes the difference between l largest values and l smallest values in set $a_i (i = 1, 2, ..., n)$, which means the moving range. $C_i^{(k)}$ is the estimated camera altitudes set at the k_{th} iteration, and C_i^{GPS} is the GPS recorded constant camera altitudes set. The moving range of the camera altitudes can represent the degree of edge deformation in reconstruction, and l is set to $n/10$ where n denotes the number of images. ω_1 and ω_2 are weighting factors to adjust the influence of two penalty terms. The first term penalizes when the estimated camera locations are far from the recorded information. Meanwhile, the second term penalizes when the edge of camera trajectories deviate from the rough recorded planar camera plane (after altitude selection). Therefore the feedback function can characterize the magnitude of the distortion error in the entire capture and is proportional to the distortion error even if the GPS accuracy is low. The relationship between k_1 and the feedback function and corresponding results are shown in Fig. 3. It is observed in Fig. 3 that the curve is similar to quadratic curve but has jagged areas which are caused by noises from uncertainties of capture and calculation. Due to the inherent ambiguity of SfM, we do not put this feedback function into the cost function of BA. Heuristic algorithms are good choices to search for the optimal radial distortion coefficient.

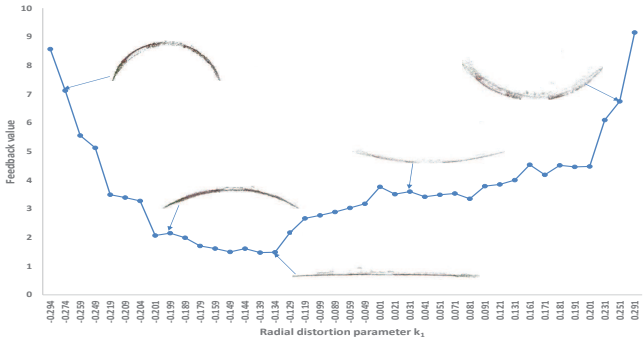


Fig. 3. Relationship between radial distortion coefficient k_1 and feedback function. The curve is obtained by changing k_1 alone and run partial BA to update feedback values. Here, the optimal value of distortion coefficient is -0.134.

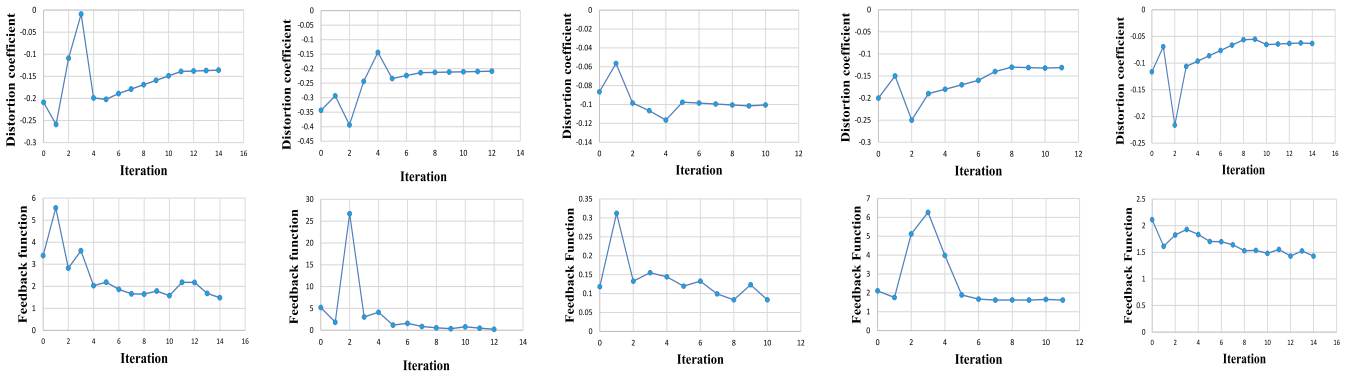
We first calculate the feedback value of the initial results obtained in Step 1. Hill climbing method [19] and simulated annealing method (SA) [20] are applied to search for the optimal distortion coefficient. If the feedback value is small, hill climbing method is applied in a small range due to its simplicity and efficiency. If the feedback value is large, SA method is applied in a relatively large range due to its superior global search ability. After each adjustment of k_1, $C_i^{(k)} (i = 1, 2, ..., n)$ are updated by invoking BA based on (2) and the feedback value is then recalculated and used to guide the next adjustment. Later experiments will show that our methods can converge efficiently according to the degree of deformation by using simple GPS information rather than prior probabilities [21] in naive baseline methods.

1) Hill climbing optimization: Hill climbing optimization is a simple yet effective method which converges fast. However, hill climbing optimization may get stuck in local optimums. To this end, hill climbing optimization is carried out multiple times and the search steps are chosen based on variation trend to overcome the uncertainties. The search is stopped if the number of iterations reach the limit or the results do not change for several iterations.

2) Simulated annealing optimization: SA optimization based on Monte Carlo iterative solution strategy is applied if the feedback value of the initial result is above a threshold. The threshold is set to 10 based on experiments. SA is known for its capability of jumping out of the local optimums. We choose a relatively high cooling rate and limit the number of iterations for efficiency. The search is stopped if the temperature reaches the limit or the results do not change for several iterations.

C. Step 3: Complete reconstruction with obtained optimal radial distortion coefficient

The entire dataset may consists of images captured under different altitudes. In step 3, we run reconstruction on the entire dataset with the optimal coefficient estimated in Step 2. And other camera intrinsics from step 2 will also be used. In other



(a) (b) (c) (d) (e)

Fig. 4. Optimization process of hill climbing method. The first row shows curves of distortion coefficient changing with iteration. The second row shows curves of feedback values changing with iteration. (a) sUAV (part, images strip). (b) Kagaru (part, images strip). (c) UAV1(part, images block). (d) sUAV (images loop). (e) Clifton (scattered images).

TABLE I
DETAILS OF DATASETS.

Dataset	Size	EXIF	GPS accuracy	k_1 Information
sUAV	313	No	5 to 10m	Yes
UAV1 [4]	147	Yes	5 to 10m	No
Kagaru [22]	360	No	2.5m	Yes
Clifton [23]	67	Yes	3m	Yes

words, images under critical configuration and the rest of the images will be reconstructed by typical SfM together in this step.

IV. EXPERIMENTS

A. Datasets

We choose four image datasets shown in Table I captured with different flight plans to validate the effectiveness of our approach. The images in each dataset are captured by the same camera except for Kagaru. Images in Kagaru are captured by two cameras, Camera 0 and Camera 1. We choose 360 images at regular intervals from Camera 0.

B. Radial distortion coefficient optimization

In this section, we choose images with similar GPS altitudes to estimate the distortion coefficient. Images captured with different flight plans are evaluated: 50 images strip and 43 images strip, 313 images loop, 147 images block and 67 scattered images. These flight plans are widely used in aerial survey.

1) Results of hill climbing optimization: Fig. 5 shows reconstructions using hill climbing optimization and reconstructions obtained with the ground truth distortion coefficient. The first two columns are results with ground truth distortion coefficient and results of VSFM [2]. The last three columns show our results from different iterations. Images in UAV1 do not have calibration information, so the ground truth result is obtained with $k1 = -0.104$ which is estimated from SA with huge

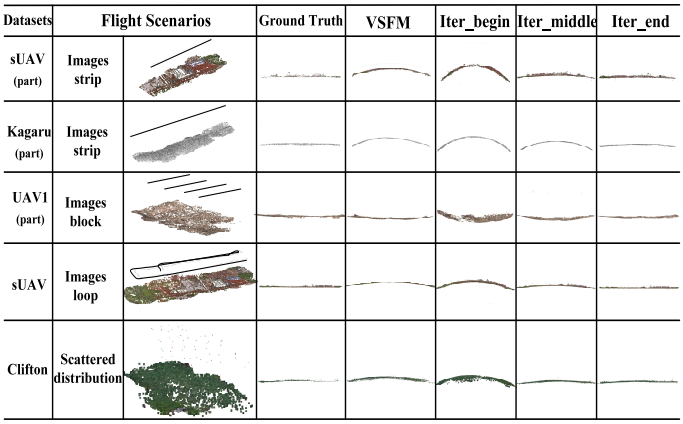
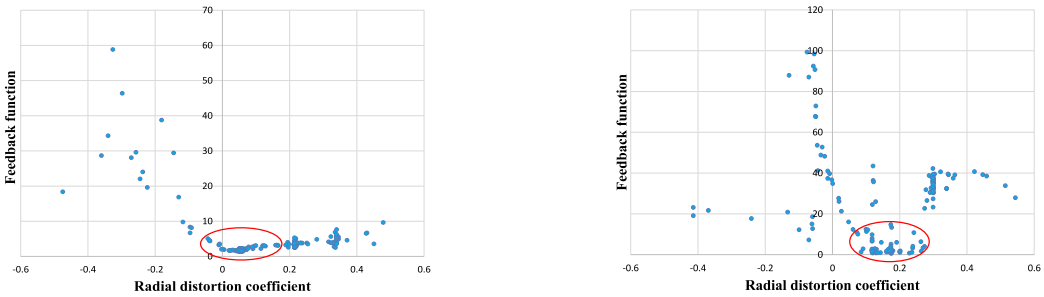


Fig. 5. Reconstructions obtained with the ground truth distortion coefficient and results of hill climbing method. Solid lines in the flight scenarios show the trajectories of the flight. The last three columns show our results from different iterations. The ground truth result of UAV1 is obtained with $k1 = -0.104$.

amount of iterations. And reconstruction under this estimated k_1 is very close to the best result in [4]. As we can see in Fig. 5, at the beginning of the optimization, reconstructions show severe doming deformations, especially in the strip scenario. The convex deformation indicates that the estimated distortion coefficient is smaller than the true value, which is partly due to that many SfM tools initialize the distortion coefficient to zero [13]. VSFM produces similar distorted results in Fig. 5. It is observed that although the reconstructions at the beginning of the optimization are not good, hill climbing method can effectively mitigate the deformation with iterations going on. Our final reconstruction results is very close to the results using the ground truth distortion coefficient.

Fig. 4 shows the optimization process of hill climbing on 5 datasets. The first and second rows show curves of distortion coefficient and feedback values changing with iteration, respectively. It is found in Fig. 4 that hill climbing can generally



(a) (b)

Fig. 6. Optimization process of SA on sUAV and Kagaru. The scattered points are testing points in the searching process. The optimal values estimated are 0.065 and 0.190, respectively. (a) Results on sUAV. (b) Results on Kagaru.

TABLE II
GROUND TRUTH DISTORTION COEFFICIENT AND ESTIMATED RESULTS FROM OUR APPROACH. COEFFICIENT OF UAV1 IS ESTIMATED FROM SA. COEFFICIENT OF KAGARU IS -0.19978 AND WE ROUND IT. BEGIN, MIDDLE AND END ARE THE RESULTS IN THE FIRST, MIDDLE AND FINAL ITERATION.

Dataset	Ground truth	Begin	Middle	End	Error
sUAV(part)	-0.134	-0.259	-0.169	-0.136	-0.002
Kagaru(part)	**-0.200**	-0.294	-0.213	-0.209	-0.009
UAV1(part)	**-0.104**	-0.057	-0.096	-0.101	0.003
sUAV	-0.134	-0.190	-0.160	-0.131	0.003
Clifton	-0.059	-0.106	-0.055	-0.063	-0.004

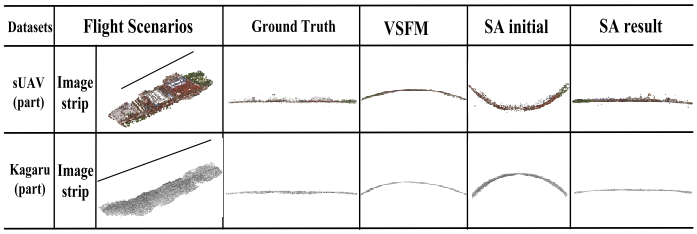


Fig. 7. Reconstruction results obtained with ground truth distortion coefficient and results of VSFM and SA. It is observed that our approach effectively eliminates the distortion error.

converge within 20 iterations when the approximate minimal feedback value is found. Table II shows the comparison between our estimated distortion coefficient and the ground truth value. It shows that hill climbing can effectively estimate the optimal distortion coefficient. The results of sUAV and Kagaru (video sequences) are worse than others, which is partly due to the lack of EXIF information which can provide better initial focal length and the error accumulation of sequences.

2) Results of simulated annealing optimization: We manually add extra radial distortions to images of sUAV and Kagaru (images strip scenario) to evaluate the performance of the SA, and the ground truth values are 0.065 and 0.190. Fig. 6 shows the optimization process of SA and Fig. 7 shows reconstructions on sUAV and Kagaru. We find that SA can generally converges within 200 iterations and our results are very close to results with ground truth distortion coefficient. On the contrary, VSFM produces obvious distorted results. Table III shows the comparison between the final three candidate coefficients of SA and the ground truth coefficient.

TABLE III
GROUND TRUTH DISTORTION COEFFICIENT COMPARED WITH THE THREE CANDIDATES ESTIMATION OF SA. WE MANUALLY ADD EXTRA RADIAL DISTORTIONS TO SUAV AND KAGARU AS INDICATED BY THE BLACK BOLD NUMBERS. ERROR IS THE ABSOLUTE ERROR OBTAINED BY COMPARING THE BEST CANDIDATE RESULT AND THE GROUND TRUTH.

Dataset	Ground truth	cand1	cand2	cand3	Error
sUAV(part)	**0.065**	0.061	0.051	0.056	-0.004
Kagaru(part)	**0.190**	0.175	0.229	0.173	-0.015

C. Complete reconstruction

Three datasets (sUAV, UAV1 and Kagaru) are chosen in complete reconstruction. Reconstruction is carried out on the entire dataset with the estimated coefficient. Fig. 8 shows a comparison among VSFM, Bundler, our approach and results reconstructed with the ground truth distortion coefficient. VSFM and Bundler produce unreasonable results marked with red boxes in Figure 8 on Kagaru where the results break into parts. And on sUAV Bundler produces obviously wrong results due to an over-estimated k_1. Although Bundler and VSFM estimate the distortion coefficients correctly on UAV1, but some camera positions from VSFM are unreasonable and some 3D points are beyond the normal boundary in Bundler. By contrast, our approach effectively eliminate the convex or concave doming deformations through optimization. Moreover, our method can adjust camera positions and 3D points to the reasonable places on UAV1. Table IV shows a comparison of feedback values from different methods. It is found in Table IV that our approach gives the lowest error.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we propose an approach to SfM with unknown radial distortions. The radial distortion coefficient estimation is optimized by taking full advantage of the auxiliary GPS

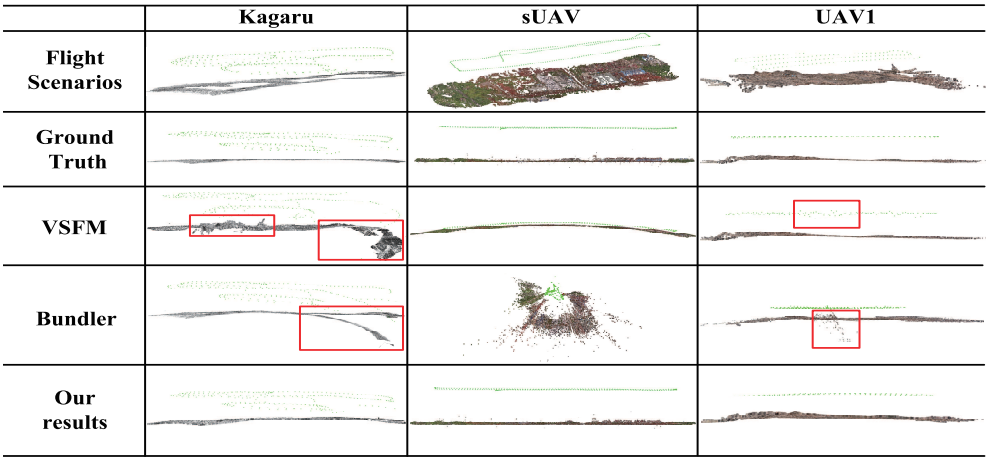


Fig. 8. Reconstructions with the ground truth distortion coefficient, VSFM, Bundler and our method. Red boxes mark the unreasonable areas.

TABLE IV
COMPARISON BETWEEN FEEDBACK VALUES OF OUR METHOD, BUNDLER AND VSFM. OUR METHOD EFFICIENTLY REDUCE THE DISTORTION ERROR AND IMPROVES THE ACCURACY OF DISTORTION ESTIMATION.

Dataset	VSFM	Bundler	Our result
Kagaru	2.12299	1.41965	**0.23234**
sUAV	2.10812	6.74955	**1.61683**
UAV1	0.12357	0.15504	**0.08345**

information. Extensive experimental results show that our approach can effectively estimate the radial distortion and reduce distorted reconstruction deformations, which are of great value in UAV applications.

ACKNOWLEDGMENT

The authors sincerely thank NVIDIA Corporation for providing high-performance graphics. This work is supported by National Natural Science Foundation of China #61503381.

REFERENCES

[1] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," *International Journal of Computer Vision*, vol. 80, no. 2, pp. 189–210, 2007.

[2] C. Wu, "Towards linear-time incremental structure from motion," in *3D Vision - 3DV 2013, 2013 International Conference on*, June 2013, pp. 127–134.

[3] D. Crandall, A. Owens, N. Snavely, and D. Huttenlocher, "Discrete-continuous optimization for large-scale structure from motion," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, June 2011, pp. 3001–3008.

[4] H. Cui, S. Shen, W. Gao, and Z. Hu, "Efficient large-scale structure from motion by fusing auxiliary imaging information," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3561–3573, Nov 2015.

[5] P. Moulon, P. Monasse, and R. Marlet, "Global fusion of relative motions for robust, accurate and scalable structure from motion," in *The IEEE International Conference on Computer Vision (ICCV)*, December 2013.

[6] N. Jiang, Z. Cui, and P. Tan, "A global linear method for camera pose registration," in *The IEEE International Conference on Computer Vision (ICCV)*, December 2013.

[7] F. S. Guo and W. Gao, "Batch reconstruction from uav images with prior information," *Zidonghua Xuebao/acta Automatica Sinica*, vol. 39, no. 6, pp. 834–845, 2013.

[8] O. Bernard and S. Souissi, "Plantwide control: the search for the self-optimizing control structure," *Journal of Process Control*, vol. 10, no. 5, pp. 487–507, 2000.

[9] S. Agarwal, N. Snavely, S. M. Seitz, and R. Szeliski, "Bundle adjustment in the large," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 29–42.

[10] M. I. Lourakis and A. A. Argyros, "Sba: A software package for generic sparse bundle adjustment," *ACM Transactions on Mathematical Software (TOMS)*, vol. 36, no. 1, p. 2, 2009.

[11] C. Wu, S. Agarwal, B. Curless, and S. M. Seitz, "Multicore bundle adjustment," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, June 2011, pp. 3057–3064.

[12] H. Cui, S. Shen, W. Gao, and Z. Hu, "Fusion of auxiliary imaging information for robust, scalable and fast 3d reconstruction," in *Computer Vision–ACCV 2014*. Springer, 2014, pp. 227–242.

[13] C. Wu, "Critical configurations for radial distortion self-calibration," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[14] L. Javernick, J. Brasington, and B. Caruso, "Modeling the topography of shallow braided rivers using structure-from-motion photogrammetry," *Geomorphology*, vol. 213, pp. 166 – 182, 2014.

[15] R. Wackrow and J. H. Chandler, "Minimising systematic error surfaces in digital elevation models using oblique convergent imagery," *The Photogrammetric Record*, vol. 26, no. 133, pp. 16–31, 2011.

[16] A. Irschara, C. Hoppe, H. Bischof, and S. Kluckner, "Efficient structure from motion with weak position and orientation priors," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. IEEE, 2011, pp. 21–28.

[17] S. Fuhrmann, F. Langguth, and M. Goesele, "Mve-a multiview reconstruction environment," in *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage (GCH)*, vol. 6, no. 7. Citeseer, 2014, p. 8.

[18] D. C. Brown, "Decentering distortion of lenses," *Photometric Engineering*, vol. 32, no. 3, pp. 444–462, 1966.

[19] S. S. Skiena, *The algorithm design manual: Text*. Springer Science & Business Media, 1998, vol. 1.

[20] S. P. Brooks and B. J. T. Morgan, "Optimization using simulated annealing," *Journal of the Royal Statistical Society*, vol. 44, no. 2, pp. 241–257, 1995.

[21] T. Sattler, C. Sweeney, and M. Pollefeys, "On sampling focal length values to solve the absolute pose problem," in *European Conference on Computer Vision*, 2014, pp. 828–843.

[22] M. Warren, D. McKinnon, H. He, A. Glover, M. Shiel, and B. Upcroft, "Large scale monocular vision-only mapping from a fixed-wing suas," in *Field and Service Robotics*. Springer, 2014, pp. 495–509.

[23] S. Gray, "Uav survey data from clifton camp (st56557330), bristol, uk," *Journal of Open Archaeology Data*, vol. 4, p. e3, 2015.