

Camera Auto-calibration for Planar Aerial Imagery, Supported by Camera Metadata

Abdullah Akay^{*†}, Hadi AliAkbarpour^{*}, Kannappan Palaniappan^{*} and Guna Seetharaman[‡]

^{*}Department of Computer Science, University of Missouri, Columbia, MO 65211 USA

[†]Department of Computer Engineering, Gebze Technical University, Kocaeli TURKEY

[‡]Advanced Computing Concepts, Naval Research Laboratory, Washington DC 20375 USA

{akaya, aliakbarpourh, palaniappan}@missouri.edu, guna@ieee.org

Abstract—UAVs in both civil and defense aviation have found numerous application areas in the last decade which resulted in sophisticated systems that extract high level information utilizing the visual data from UAVs. Camera auto-calibration is always the first step if the intrinsic parameters of the cameras are not available. However, this process is not trivial as the aerial imagery mostly contains planar scenes which constitute a degenerate condition for conventional methods. In this paper, we propose a hybrid approach which incorporates circular point and camera position constraints as a single optimization term to automatically calibrate the camera of the UAV on planar scenes. The experimental results show that our proposed hybrid method is more robust and accurate than the conventional counterparts.

Index Terms—Auto-calibration, Planar Scene Reconstruction, Metadata Recovery, Aerial Imagery

I. INTRODUCTION

The use of Unmanned Aerial Vehicles (UAV) has been getting widespread in both civil and defense areas in the last decade. As a result of this, the smart systems for analysis of aerial images from UAVs are getting significant importance [1]–[9]. Such systems require some fundamental components like calibration. However, the standard computer vision camera calibration methods aren't suitable for all cases in aerial imagery; particularly when the focal length abruptly changes along the video sequence because of zooming in or out [10]–[16].

In order to cope with these obstructions, auto-calibration techniques can be considered [17]. Auto-calibration is important because it can handle the problem of varying internal camera parameters without a need to use any artificial geometric pattern in the scene. Nonetheless, the traditional auto-calibration methods have some limitations on aerial imagery as the scenes in that kind of scenarios are usually planar which causes a degenerate case for most of the existing techniques [18]–[22]. A method for auto-calibration on planar scenes is proposed in [23]. Their method tries to minimize a cost function to estimate circular points of the first image and hence the intrinsic matrix K . It works well on synthetic data. However, in the real case it mostly produces spurious solution because of the excessive number of parameters to be recovered and noise. On the other hand, if some constraints on those unknown parameters are used, it can be possible to come up with a reliable solution.

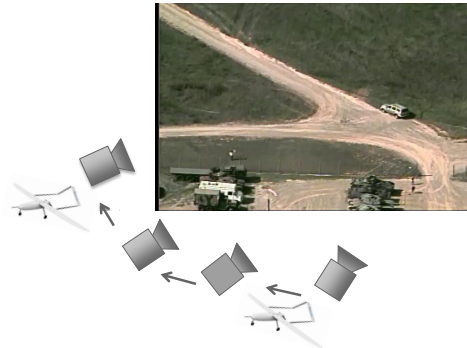


Fig. 1: We use images from different viewpoints of a planar scene with GPS data.

In this paper, we developed a constrained system for auto-calibration from aerial images and followed by a 3D camera pose estimation. We get images of a planar scene from different viewpoints with GPS data for each frame (Fig. 1). We used GPS position of the UAV as a constraint on the unknowns to suppress spurious solutions. The unknown true scale is determined by a procedure in which the dimensions of a real object (for example a car) are compared to the dimensions of a reconstructed one.

Our solution is a hybrid approach that takes advantage of camera positions along with two complex conjugate points known as *circular points* which are basically intersections of the absolute conic and the image plane [18]. This approach performs well on planar scenes. Moreover, it is lightweight compared to conventional metadata estimation algorithms that employ a complete 3D reconstruction because of two reasons. Firstly, we require only 4 or more points correspondences for each consecutive frame. Second, our method has relatively low computational cost which gives a significant advantage especially when repetitive calibrations are required because of the varying internal parameters along a flight.

We examined the proposed system on controlled and real datasets. The experimental results show that our system is able to effectively produce reliable solutions for the problem of auto-calibration and 3D camera pose estimation using aerial imagery.

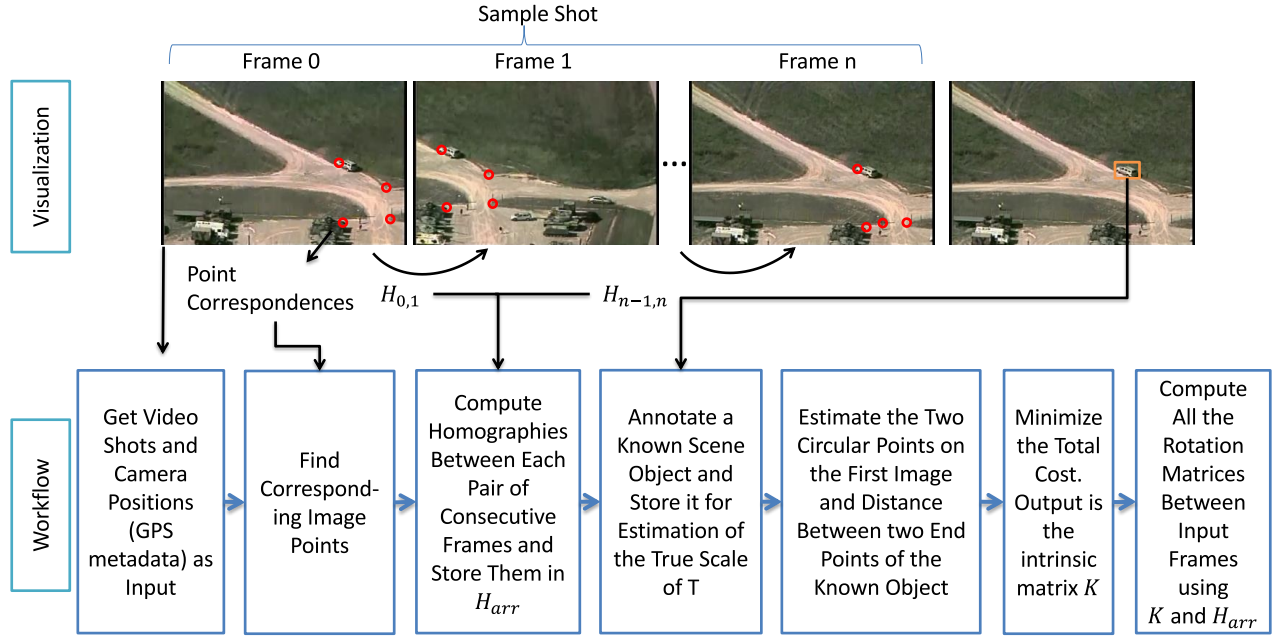


Fig. 2: Overview of our approach.

A. Our Contributions

- We propose a hybrid approach that makes the use of circular points and camera positions to increase robustness of auto-calibration.
- A relatively lightweight method in terms of computational cost and required input that works well on planar scenes in aerial imagery.

II. THE METHOD

We first describe some concepts of auto-calibration that we utilized in the proposed method.

A. Calibration with Circular Points

The circular points of an image can be computed by following some certain steps. Let's say H is a 3×3 homography matrix from A to B (Fig. 3). Note that, A is an 2D image of a 3D square object and B is a 2D square. Then, $h_1 \pm i h_2$ is the circular points of that image where, h_1 , h_2 and h_3 are the column vectors of H .

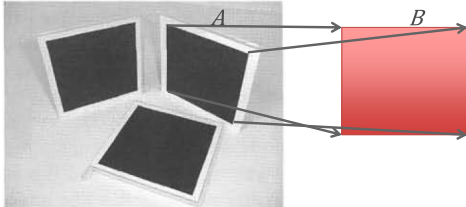


Fig. 3: Computation of circular points.

Once we get 3 distinct circular points for an image, we can compute the image of absolute conic by minimizing (1).

$$(h_1 \pm i h_2)^T \omega (h_1 \pm i h_2) = 0 \quad (1)$$

where, ω is the image of absolute conic.

Determining ω is particularly important as it is known that $\omega = (K K^T)^{-1}$. Therefore, the intrinsic matrix K can be recovered by applying Cholesky Decomposition to ω . However, this method is only valid if the circular points are known. In a real scenario, there is no special patterns or objects in the scene allowing us to compute circular points directly.

B. Auto-calibration with Circular Points

Circular points are transferred from frame to frame by inter-image homographies. Theoretically, if there are 3 images available, 4 parameters of circular point c_j and 1 parameter of K can be recovered by minimizing (2).

$$(H^i c_j)^T \omega (H^i c_j) = 0, \quad i = 1, \dots, m \text{ and } j = 1, 2 \quad (2)$$

where $H^i c_j$ is the circular point in image i , m is scene count, H^i is inter-image homography. The reason why there is only one parameter of K is that we only recover focal length as we assume principle point is the center of the image plane and there is no skew. The 5 unknown parameters can be iteratively estimated using any kind of meta-heuristic algorithms.

This basic approach works pretty well on synthetic data which have small amount of noise. However, for the real one, it requires 40-50 frames taken from a set of viewpoints having a rotational variance of 30-40 degrees in each axis which is not generally the case. After all, it works with planar scene points, but it is quite sensitive. As a result of this, it estimates K poorly in general. We will devise a new hybrid method to compensate these weaknesses which will be described in Subsection II-C1.

Algorithm 1 Estimation of K

Input: A video shot and position (GPS) data of the camera
Output: Final K and camera poses

Initialization: Compute point correspondences and inter-image homographies for each consecutive frame (H_{arr})

- 1: **while** max iteration count is not reached **do**
 - 2: Guess/Update K
 - 3: Guess/Update two complex circular points (CP) on the first image
 - 4: $CP_{Ti} := H_i * CP_i$ {Transform the CP to all consecutive frames using H_{arr} }
 - 5: Compute how well all the circular points fit in current K (the first term of (3)) and use it as circular point cost $Cost_{CP}$
 - 6: Compute rotation matrix R for each frame by decomposing H_i and K
 - 7: Reconstruct two known points using K , R and GT positions
 - 8: Compute true scale using given two points
 - 9: Compute distance between two points and use it as distance cost $Cost_{dist}$ (the second term of (3))
 - 10: Merge $Cost_{CP}$ and $Cost_{dist}$ into total cost and store the arguments with the minimum cost
 - 11: **end while**
-

C. Metadata Recovery

Our aim is to recover camera parameters and poses for each frame in a video shot. We first get shots of an input video using [24]. We assume that internal camera parameters remain the same within a shot, but can change between different shots. We proposed two different methods to recover metadata for different scenarios.

1) *Metadata recovery using known camera positions:* In order to increase the robustness of the method described in Section II-B, we used position data of the camera to put an extra constraint to the system. Firstly, we compute point correspondences between frames using SIFT (Fig. 2). The homography matrices between each consecutive frame are computed using these points. The intrinsic matrix K is guessed randomly at initialization phase. We try to minimize two different cost terms (3). The first one is the cost of transferring circular points from first frame to the last frame which is represented by the first term of (3). This is basically carried out by the approach described in Section II-B. For the second term, extrinsic parameter R is computed using K and H^i with homography decomposition algorithm of [25]. Since we assume that the translation vector t is known, we can reconstruct scene points up to a scale. A scene objects is manually annotated to find the true scale of the reconstruction. After two end points of the annotated object are reconstructed using estimated K , R and known t , the distance between the ground truth length and computed length of the object (e.g. length of a car) is computed which is represented by the second term of (3). The circular point cost and the distance cost are merged to constitute a hybrid cost function which is minimized

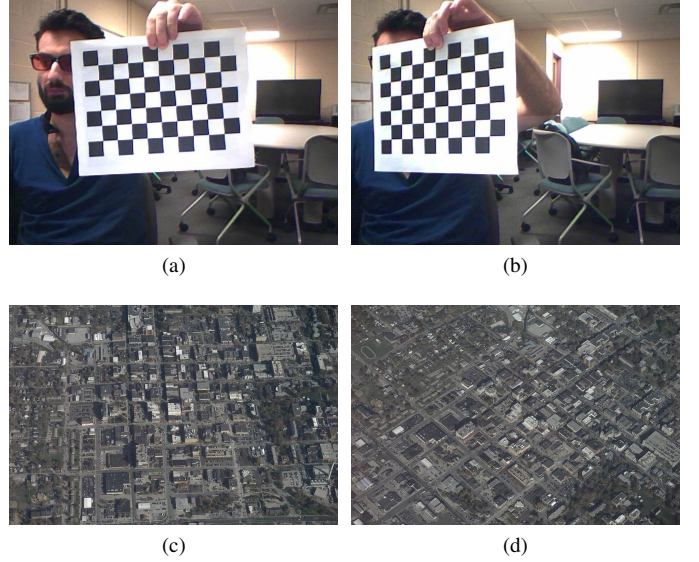


Fig. 4: Some examples from our datasets. (a), (b) calibration pattern. (c), (d) city of Columbia

iteratively (3).

$$\operatorname{argmin} \| (CP_{Ti} \omega CP_{Ti})^2 + \alpha(T_{REC} - T_{GT})^2 \| \quad (3)$$

where $CP_{Ti} = H^i c_j$ is circular points of each frame. While T_{REC} is the reconstructed length of the scene object, T_{GT} is the ground truth length of that object. α is weighting parameter which is set to 0.6 empirically. The range of α is between 0-1. One can safely pick another value between the range to change the balance between the constraint terms.

We used a search technique similar to the genetic algorithms followed by a Levenberg-Marquart minimization to fine-tune the results of our method in each iteration. The details of the described approach is presented in Algorithm 1.

2) *Metadata recovery using parallel lines:* We developed another solution for the scenario in which the GPS data are unknown. In this case, we ask the user to provide two pair of parallel lines in the first image (for example edges of a pair of roads). We use this information to put a constraint on circular point equation (2). Originally, circular points have 4 DOF. However, using the parallel lines it decreases to 2 DOF which makes it easy to estimate the intrinsic matrix K using search algorithms. The only difference between this approach and the one described in Section II-C1 is that we don't use any camera position constraints. The other parts mostly remain the same. The output of the both methods are quite similar.

III. EXPERIMENTAL RESULTS

We conducted various experiments on different datasets.

A. Datasets

We prepared 3 different datasets to assess our method thoroughly (Fig. 4).

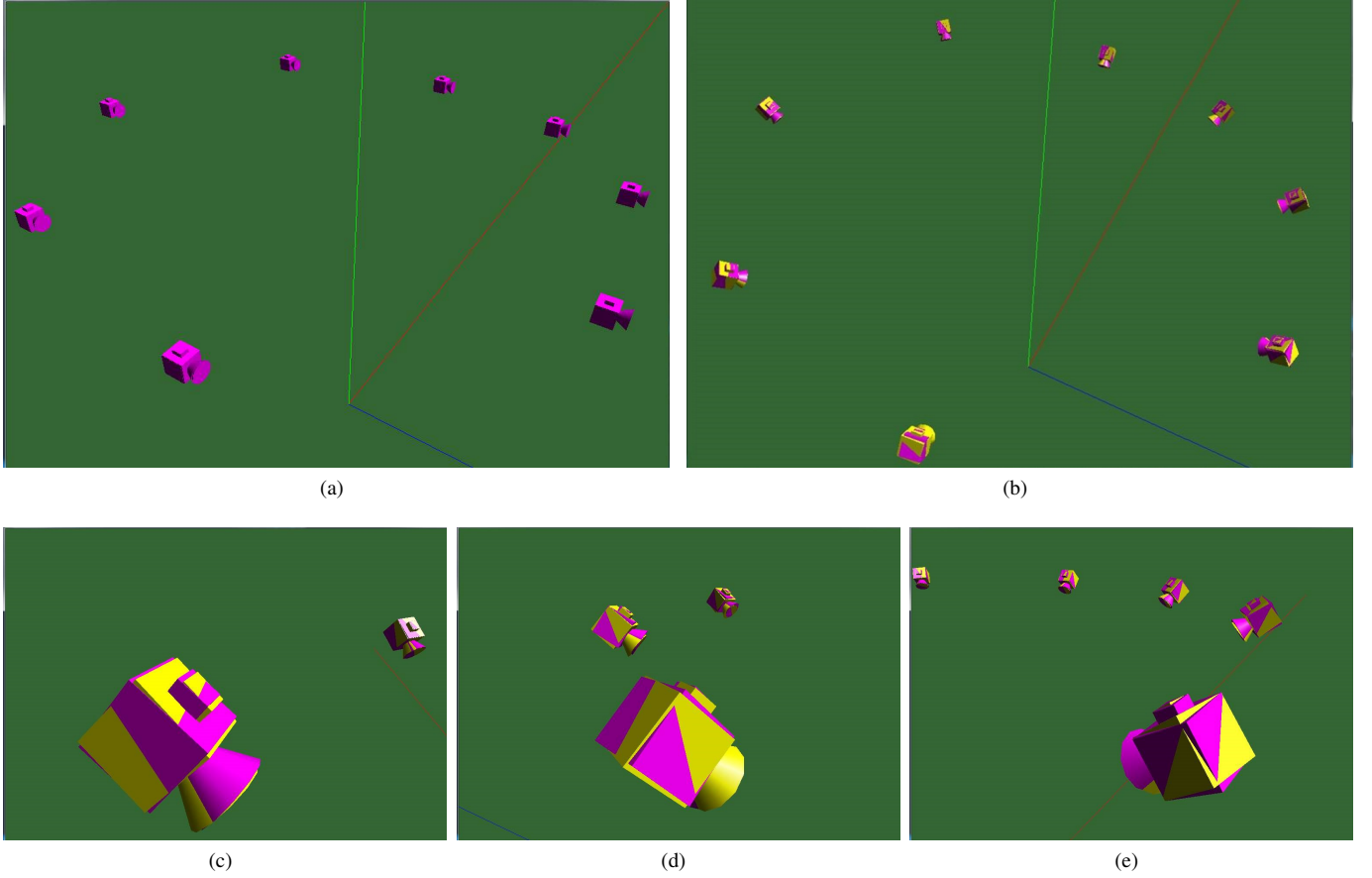


Fig. 5: A bunch of reconstructed camera poses. (a) initial poses, (b) estimated camera poses. (c), (d), (e) detailed snapshots of individual camera poses. Note that, the cameras represented in yellow are ground truth poses and the ones represented in purple are estimated poses.

- Synthetic: This set consists of synthetic images of a rectangle with different rotations and translations.
- Calibration Pattern: We took images of standard calibration pattern of OpenCV library from different viewpoints.
- Columbia: Aerial images of city of Columbia, MO taken by an UAV. The UAV flew around the city in a circular manner viewing the city center.

B. Experiment Types

We performed 3 types of experiment to measure the accuracy of the procedures of estimating matrix K .

1) *Circular Points (CPA)*: In this case, we estimated K using only circular point constraint (2) which can be seen as a baseline method.

2) *Estimating K using only position data (KPD)*: We searched for K using a meta-heuristic algorithm and performed limited 3D reconstruction using K , R and given t . Then, we compute the distance cost (second term of 3) and use only it as cost function.

3) *Circular Points + Position Data (CP+PD)*: In the final configuration which is actually proposed solution, we incorporated both constraints into a single one (3). Result of each type

of experiments on different datasets are presented in Table I. In each column labeled as **E1**, average percentage of focal length error in pixels can be seen. Note that, lower ones are better.

C. Computational Evaluation

We also measured the computational performance of our method. In the columns of Table I corresponding to label **E2**, average iteration count for each experiment configuration can be seen.

A single iteration of execution doesn't take the same amount of time for each of the configurations. Hence, we also determined the computational cost of the each experiment type in seconds in the columns of Table I labeled as **E3**. None of the implementations we used in experiments are optimized. We used a laptop with a Core 2 Duo processor and a 3 GB of memory to run the experiments. Again, lower ones are better.

D. Camera Poses

Some qualitative results of the proposed method are presented in Fig. 5. While, the initial poses of the cameras are shown in Fig. 5a, the final estimated poses are given in Fig. 5b.

TABLE I: Quantitative results of the proposed method

Method	Synthetic			Calibration Pattern			Columbia		
	E1	E2	E3	E1	E2	E3	E1	E2	E3
CPA	2.31	55	0.6	16.92	156	1.5	- ^a		
KPD	1.85	12	1.1	15.08	33	3.2	12.82	49	4.8
CP+PD	0.92	8	1.4	3.07	32	3.8	2.56	53	5.9

^aDoesn't converge.

Some detailed examples of our results can be seen at the rest of the sub-figures. The size of all the camera representations are scaled up for visualizations proposes. In the real case, the distance between two camera poses from two ends is about 3 kilometers. Finally, we give the average rotational difference between the ground truth poses and the reconstructed ones for each axis in Table II to demonstrate the accuracy of our pose recovery results.

TABLE II: Average angle error in degrees

Axis	Angle Error
X	3.29
Y	4.36
Z	3.03

IV. CONCLUSIONS

In this paper, we proposed a novel approach for recovering partially missing metadata in aerial imagery. Various experiments show that the proposed hybrid approach is more robust and accurate than the conventional counterparts. Our method is computationally low cost and efficient for runtime applications. The proposed method can be extended to recover other types of metadata such as position that may eliminate the need of GPS module in extreme cases.

ACKNOWLEDGMENT

This work is sponsored by The Scientific and Technological Research Council of Turkey (TUBITAK) and Gebze Technical University, Turkey.

REFERENCES

- [1] H. Aliakbarpour, K. Palaniappan, and G. Seetharaman, "Oblique aerial image georegistration and efficient camera pose refinement without correspondence-based homographies."
- [2] K. Palaniappan, M. Poostchi, H. Aliakbarpour, R. Viguier, J. Fraser, F. Bunyak, A. Basharat, S. Suddarth, E. Blasch, R. Rao, and G. Seetharaman, "Moving object detection for vehicle tracking in wide area motion imagery using 4d filtering," in *Proc. IEEE International Conference on Pattern Recognition (ICPR)*. [Online]. Available: <http://icpr2016.org/>
- [3] R. Aktar, V. B. S. Prasath, H. Aliakbarpour, U. Sampathkumar, G. Seetharaman, and K. Palaniappan, "Video haze removal and poisson blending based mini-mosaics for wide area motion imagery," in *Proc. IEEE Applied Imagery Pattern Recognition (AIPR)*.
- [4] H. AliAkbarpour, K. Palaniappan, and G. Seetharaman, "Fast structure from motion for sequential and wide area motion imagery," in *Proc. IEEE International Conference on Computer Vision Workshop (ICCVW) Video Summarization for Large-scale Analytics Workshop*. [Online]. Available: <https://sites.google.com/site/lsvsicc2015/video-summarization-for-large-scale-analytics-workshop>
- [5] H. Aliakbarpour, K. Palaniappan, and G. Seetharaman, "Robust camera pose refinement and rapid SfM for multiview aerial imagery without RANSAC," vol. 12, no. 11, pp. 2203–2207. [Online]. Available: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7185357>
- [6] M. Shahbazi, J. Thau, and P. Mnard, "Recent applications of unmanned aerial imagery in natural resource management," vol. 51, no. 4, pp. 339–365.
- [7] Y. Su, Q. Guo, D. L. Fry, B. M. Collins, M. Kelly, J. P. Flanagan, and J. J. Battles, "A vegetation mapping strategy for conifer forests by combining airborne lidar data and aerial imagery," vol. 42, no. 1, pp. 1–15.
- [8] S. Workman, R. Souvenir, and N. Jacobs, "Wide-area image geolocalization with aerial reference imagery," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3961–3969.
- [9] B. O. Abayowa, A. Yilmaz, and R. C. Hardie, "Automatic registration of optical aerial imagery to a LiDAR point cloud for generation of city models," vol. 106, pp. 68–81.
- [10] I. Colomina and P. Molina, "Unmanned aerial systems for photogrammetry and remote sensing: A review," vol. 92, pp. 79–97.
- [11] J. A. Goncalves and R. Henriques, "UAV photogrammetry for topographic monitoring of coastal areas," vol. 104, pp. 101–111.
- [12] R. Kumar, H. Sawhney, S. Samarasekera, S. Hsu, H. Tao, Y. Guo, K. Hanna, A. Pope, R. Wildes, and D. Hirvonen, "Aerial video surveillance and exploitation," vol. 89, no. 10, pp. 1518–1539.
- [13] C.-C. Lin, S. Pankanti, G. Ashour, D. Porat, and J. R. Smith, "Moving camera analytics: Emerging scenarios, challenges, and applications," vol. 59, no. 2, pp. 5–1.
- [14] S. Siebert and J. Teizer, "Mobile 3d mapping for surveying earthwork projects using an unmanned aerial vehicle (UAV) system," vol. 41, pp. 1–14.
- [15] K. P. Valavanis and G. J. Vachtsevanos, *Handbook of unmanned aerial vehicles*. Springer Publishing Company, Incorporated.
- [16] A. Banaszek, A. Zarnowski, A. Cellmer, and S. Banaszek, "Application of new technology data acquisition using aerial (UAV) digital images for the needs of urban revitalization."
- [17] P. J. Zarco-Tejada, R. Diaz-Varela, V. Angileri, and P. Loudjani, "Tree height quantification using very high resolution imagery acquired from an unmanned aerial vehicle (UAV) and automatic 3d photo-reconstruction methods," vol. 55, pp. 89–99.
- [18] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision second edition."
- [19] D. Liebowitz and A. Zisserman, "Combining scene and auto-calibration constraints," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 1. IEEE, pp. 293–300.
- [20] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank, "Camera self-calibration: Theory and experiments," in *European conference on computer vision*. Springer, pp. 321–334.
- [21] R. I. Hartley, "Kruppa's equations derived from the fundamental matrix," vol. 19, no. 2, pp. 133–135.
- [22] Q.-T. Luong and O. D. Faugeras, "Self-calibration of a moving camera from point correspondences and fundamental matrices," vol. 22, no. 3, pp. 261–289.
- [23] B. Triggs, "Autocalibration from planar scenes," in *European conference on computer vision*. Springer, pp. 89–105.
- [24] R. Viguier, C. C. Lin, H. AliAkbarpour, F. Bunyak, S. Pankanti, G. Seetharaman, and K. Palaniappan, "Automatic video content summarization using geospatial mosaics of aerial imagery." IEEE, pp. 249–253. [Online]. Available: <http://ieeexplore.ieee.org/document/7442335/>
- [25] E. Malis and M. Vargas, "Deeper understanding of the homography decomposition for vision-based control."