

SECOND ITERATION OF PHOTOGRAMMETRIC PIPELINE TO ENHANCE THE ACCURACY OF IMAGE POSE ESTIMATION

NGUYEN Truong Giang^a, Marc Pierrot-Deseilligny^b, Jean-Michael Muller^b, Christian Thom^b

^a UPE - University Paris-Est - Truong-Giang.Nguyen@ign.fr

^b IGN - Institut national de l'information géographique et forestière - first_name.last_name@ign.fr

Commission II, WG II/4

KEY WORDS: tie points extraction, image matching, interest points detection, orientation accuracy enhancement

ABSTRACT:

In classical photogrammetric processing pipeline, the automatic tie point extraction plays a key role in the quality of achieved results. The image tie points are crucial to pose estimation and have a significant influence on the precision of calculated orientation parameters. Therefore, both relative and absolute orientations of the 3D model can be affected. By improving the precision of image tie point measurement, one can enhance the quality of image orientation. The quality of image tie points is under the influence of several factors such as the multiplicity, the measurement precision and the distribution in 2D images as well as in 3D scenes. In complex acquisition scenarios such as indoor applications and oblique aerial images, tie point extraction is limited while only image information can be exploited. Hence, we propose here a method which improves the precision of pose estimation in complex scenarios by adding a second iteration to the classical processing pipeline. The result of a first iteration is used as *a priori* information to guide the extraction of new tie points with better quality. Evaluated with multiple case studies, the proposed method shows its validity and its high potential for precision improvement.

INTRODUCTION

Considered as the first step of digital photogrammetric processing, tie point extraction provides main observations for pose estimation (bundle block adjustment). There exist two common approaches of tie point detection for multiple images. First approach (image based) uses local search and match algorithm such as Correlation and Least Square Matching. Second approach (key point based) uses key points detection and matching over multiple images with descriptor.

The research of [Shragai et al., 2011] shows that image based method gives a good result when we know the *a priori* information of camera orientation while the key point based method adapts to general cases. Key point based tie points extraction method like SIFT [Lowe, 2004], SURF [Bay et al., 2008]... proved their robustness and reliability in various applications, from aerial to terrestrial scenarios [Lingua et al., 2009]. Tie points measured by these methods is used to obtain a reasonable photogrammetric production.

Though this approach is operational in many cases, occasionally, especially in the presence of large depth variety like complex interior terrestrial scenario, perspective deformations between neighbor images reduce its performance, limit the precision and the accuracy of pose estimation. There exist approaches which overcome the problem of high geometrical distortion between images like A-SIFT [Morel and Yu, 2009], but at the price of sacrificing the computation time and at the risk of a large number of false matches.

When it comes to fine metrology photogrammetric applications, tie points extracted only with image information may not be precise enough for pose estimation. In photogrammetric processing, the quality of tie points includes the measurement, the multiplicity and the homogeneity of the distribution in 2D images as

well as in 3D scenes. To extract tie points with enhanced quality for photogrammetric use, we propose a method that combines the information of 3D scenes and 2D images. The information of 3D scenes is obtained by a first photogrammetric processing which predicts and corrects the perspective effect between images before extracting tie points. Finally, a second bundle block adjustment using these tie points of high quality will improve significantly the precision of pose estimation.

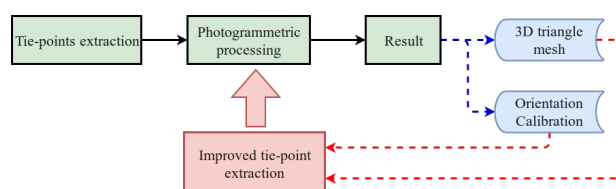


Figure 1. A second iteration based on tie points of high quality extracted with a first processing result.

METHOD

Our algorithm uses as input an approximate 3D triangle mesh and the orientation of each image obtained from a first photogrammetric processing. The algorithm being applied to each triangle in the mesh consists of 3 steps:

1. Selection of images.
2. Detection of interest points on affine invariant region.
3. Matching by correlation.

Each step is presented below in details.

Images selection

For each 3D triangle in the mesh, a set of suitable images must be selected. Firstly, 3D triangle must be visible in the field-of-view of image. A simple filter ZBuffer is used to check this condition. Then, a unique master image for this 3D triangle is selected on the basis of an optimal resolution condition. This master image is chosen to maximize the resolution of observation in all directions within the plane of the mesh.

The optimal resolution condition can be modeled by the deformation of the ellipse tensor between the plane of the triangle in 3D scene and in 2D image which is momentarily back-projected. Due to the perspective deformation, a circle will be transformed into an ellipse depending on the direction of observation. Therefore, we use here a prudent strategy by defining the minimal resolution obtained among all possible directions in image. Each 3D triangle defines a 3D plane that is related to its back-projections in 2D images by a homography transformation. Because the size of the triangle is small, this relation can be approximated by an affine transformation. To estimate the parameters of the affine transformation for an image, the triangle's vertices in 3D scene and 2D image are used.

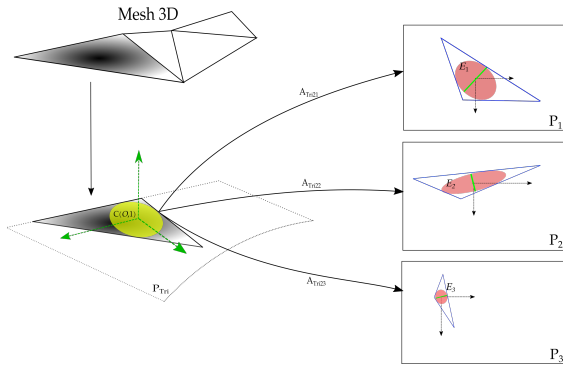


Figure 2. Master image selection by optimal resolution condition.

Note that P_n is the plane of image n , P_{Tri} is the plane of the considered 3D triangle. $C(O, 1)$ is the unit circle on the plane P_{Tri} and inside the 3D triangle. E_n is the ellipse transformed from circle $C(O, 1)$ on the plane P_{Tri} to plane image P_n by an affine transformation A_{Tri2n} as shown below:

$$A_{Tri2n} = \begin{bmatrix} A_{x_n} & B_{x_n} \\ A_{y_n} & B_{y_n} \end{bmatrix} = [\vec{u}_n \quad \vec{v}_n] \quad (1)$$

For each point $(\cos(\theta), \sin(\theta))$ on circle C , we have its correspondent point on ellipse E_n by affine transformation A_{Tri2n} :

$$\begin{aligned} A_{Tri2n} * \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} &= [\vec{u}_n \quad \vec{v}_n] \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \\ &= \vec{u}_n \cos \theta + \vec{v}_n \sin \theta = V_n(\theta) \end{aligned} \quad (2)$$

The value $\min(V_n(\theta))$ represents the minimal possible resolution obtained through all directions in image n of circle $C(O, 1)$.

$$\min(\|V_n(\theta)\|^2) = \frac{u_n^2 + v_n^2 - \sqrt{(u_n^2 - v_n^2) + 4(uv)^2}}{2} \quad (3)$$

Hence, the master image is the maximum of all possible minimal resolutions through all images: $\min(\|V_n(\theta)\|^2)$.

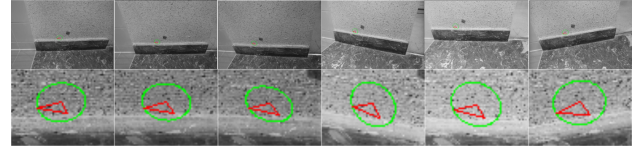


Figure 3. Different ellipses given by different directions of observation. The image on the right is selected as the master image for the concerned 3D triangle.

Finally, after selecting master image for each 3D triangle, we have a partition of the scene between different master images. The set of second images is selected with the optimal resolution condition applied to master image and other images.

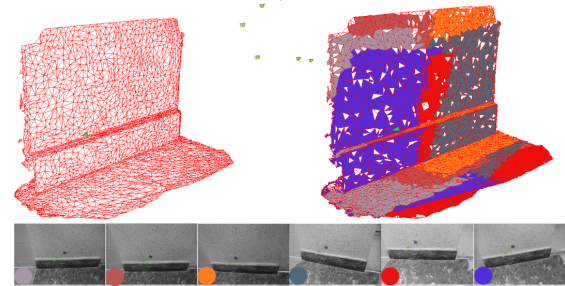


Figure 4. Partition of the scene between master images. Top-left: 3D mesh, top-right: each color represents one master image, bottom: master images with their color represented in scene.

Detection of interest points

After selection of images, we have a mesh composed of a set of 3D triangles, each containing a master and a set of slave (secondary) images. A detection of interest points is then performed on region of interest (ROI) for each master-secondary image pair in the set. Each 3D triangle is back-projected to the images to form a correspondent triangle ROI. The ROI in the secondary image is then normalized to master image geometry by an affine transformation estimated from known corresponding coordinates of back-projected triangles. Then, a detection of interest points is performed on these affine invariant ROIs.

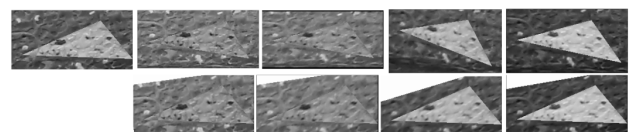


Figure 5. Upper-line: Correspondent ROIs on different images, left: ROI of master image. Lower-line: normalized ROI to master image geometry.

To optimize the matching by correlation, interest points are detected with 3 conditions:

- Local maxima points (minimum/maximum) as candidates.
- Low-contrast filter.
- Non-repetitive with adjacent patterns filter.

2.2.1 Low-contrast filter: First of all, local maxima points are selected as candidate points. Then, low-contrast points are eliminated by a filter inspired by FAST [Rosten, 2006]. For each candidate point whose intensity value is denoted p , we examine its neighboring points ($1...n$) with intensity values ($p_1...p_n$) respectively. The neighboring points of a candidate point are the pixels on a circle of radius R around it. Hence, a configuration of $n = 12$ or $n = 24$ neighboring points corresponds to $R = 3$ or $R = 4$, respectively. With an appropriate threshold value t , a neighboring point i is valid if:

$$|p - p_i| > t \quad (4)$$

By examining all neighboring points, a candidate point is considered as contrast if:

- 75% of neighboring points are valid.
- 60% of neighboring points are valid and contiguous.

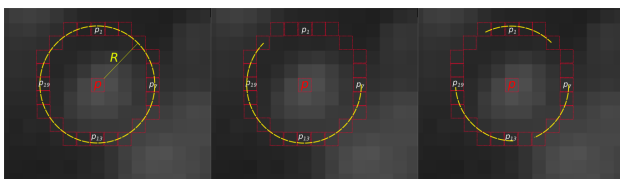


Figure 6. Low-contrast filter. Left: configuration with $R = 4$, middle: 75% of contiguous neighboring points, right: 65% of non-contiguous neighboring points

2.2.2 Non-repetitive with adjacent patterns filter: For a matching with image correlation method, a repetitive or linear pattern can cause a "false positive" error. The last filter checks if a candidate point has a risk of repetitive patterns with its adjacent neighboring points. By correlating between the image patch centered on candidate points and all-around patches, ambiguous points will be eliminated if a high-score correlation is found in any direction.

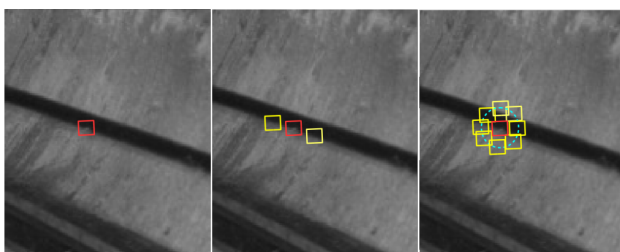


Figure 7. Non-repetitive filter. Left and middle: Risk of false match with linear pattern. Right: Verify the patch of one candidate (red) with patches around (yellow).

The configuration of patches around in non-repetitive filter is as same as low-contrast filter. An appropriate radius will be selected to form a circle of neighboring patches around candidate point.

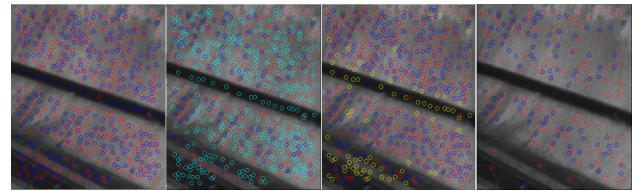


Figure 8. Detection of interest points: From the left: Local maxima points (minimum/maximum), Low-contrast points (cyan), risk of repetitive points (yellow), final points.

Matching by correlation

Since we work on ROIs of invariant geometry (all ROIs of secondary images are normalized to the geometry of the master image), a Zero Mean Normalized Cross-Correlation (ZNCC) is selected as a matching method. For each interest point in master image, we search for a point with highest correlation in the secondary images. First of all, the potential matching region is defined as a circle around interest points of master image. Then, possible match candidates are selected as interest points in secondary images which have the same characteristic (maximum/minimum) with considered point in master image. The correlator for each couple of possible match operates on three scales to speed up the process:

- n -times downsampled without interpolation (one out of n pixels is taken into account).
- nominal image scale.
- sub-pixel up to 0.01 of a pixel scale.

To remove matches of low reliability, an independent threshold value is applied at each of these three scales.

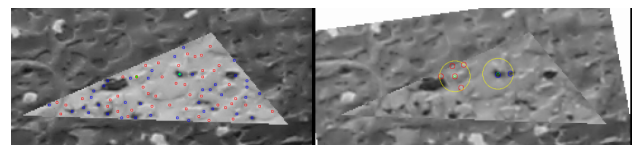


Figure 9. Potential matching region (yellow circle in secondary image ROI - right) and possible match candidates (same characteristic point: max (red), min (blue)) correspondent with considered interest point in master image (left- green)

After matching all interest points in the ROI of master image, a spatial filter is applied to homogenize tie point distribution.

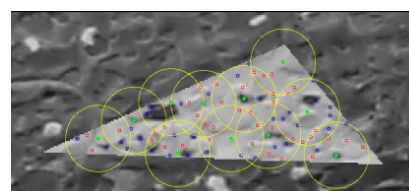


Figure 10. Right: Spatial filter (yellow) keeps only 1 matched point (green) in circle.

EXPERIMENTS AND RESULTS

The proposed method aims to re-extract high quality tie points by using information from the result of first photogrammetric processing to improve the precision of pose estimation by a second bundle block adjustment. To evaluate the performance of the method, we use two different real-case studies. For each case, a classical photogrammetric processing operates firstly to get input data for the proposed method. Then, improved tie points are extracted by the proposed method and a second bundle block adjustment is performed on these new tie points. The precision of pose estimation will be calculated with ground control points. Tie point projection error will be evaluated as well. All photogrammetric processings are realized with MicMac [Deseilligny and Clery, 2011] - a free open source project for photogrammetry.

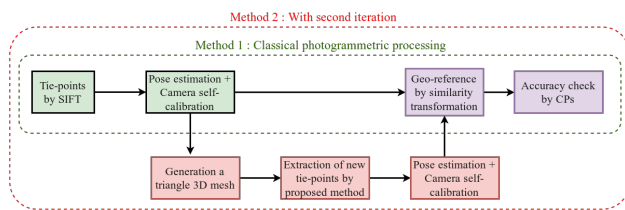


Figure 11. Two photogrammetric processing chains used in experiment.

Two photogrammetric processing chains are shown in Figure 11:

- *Method 1*: Classical photogrammetric processing.
- *Method 2*: Photogrammetric processing with second iteration.

In all cases, tie points are the unique measurement for bundle block adjustment. The 3D model is geo-referenced with a similarity transformation estimated with CPs¹. Two experiments on real case studies are described below.

3.0.1 Case 1 - a terrestrial scene This is a terrestrial acquisition of a polygon calibration structure. It contains 56 images, 9 GCPs² and 11 CPs. Images are acquired in ENSG³ with a commercially available camera Sony RX-1.

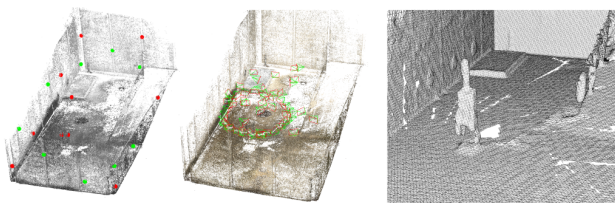


Figure 12. The terrestrial acquisition (Case 1). From the left: Distribution of GCPs (green) & CPs (red), acquisition protocol, initial mesh (from result of method 1) generated by MicMac.

The comparison of accuracy on CPs between method 1 and method 2 is given in Figure 13 and it shows a significant enhancement. The residuals (euclidean distance) decrease on 10/11

¹Control Point

²Ground Control Point

³Ecole nationale des sciences géographiques - France

of CPs. The average residual decreases by a ratio of 2.42, from 1.55mm down to 0.64mm. Accordingly, average image residual (reprojection error) is also decreased from 0.57 to 0.36 pixel Figure 14.

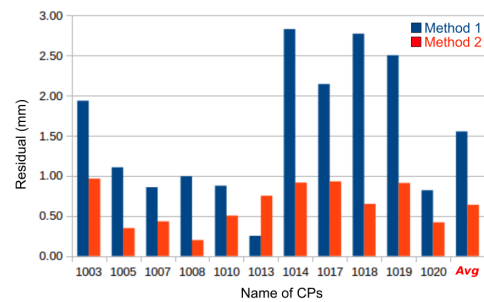


Figure 13. Accuracy check on CPs - Case 1

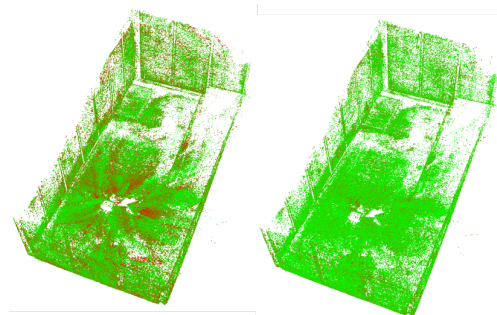


Figure 14. Image residual (reprojection error) estimated from camera orientation given by method 1 (left) and method 2 (right). Residual of tie points is represented in different colors (from 0 (pure green) to 1 (pure red))

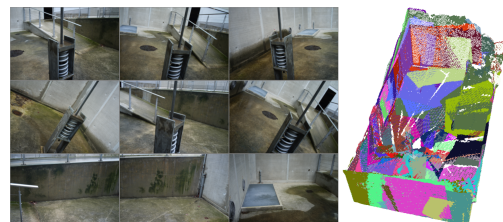


Figure 15. Partrition of scene between master images in Case 1. Left: some images, right: each color represents a selected part of the scene on one master image.

3.0.2 Case 2 - an UAV acquisition scene This is a simple UAV acquisition containing 73 images, 5 GCPs and 12 CPs. Images are acquired by Vinci-Construction-Terrassement UAV with IGN photogrammetric camera CamLight [Martin et al., 2014]. Details about the distribution of CPs, acquisition protocol and initial mesh are given by classical photogrammetric pipeline (method 1) as shown in Figure 16.

Same as Case 1, new tie points show a notable improvement in the accuracy check computed on CPs. The residuals (euclidean distance) decrease on 11/12 of CPs. The average residual decreases by a ratio of 1.75, from 30.33mm down to 17.54mm.

For all test cases, details of accuracy check along axis X, Y, Z on each CP are given in Figure 4.



Figure 16. The UAV acquisition (Case 2). From the left: Distribution of GCPs (green) & CPs (red), acquisition protocol, initial mesh (from result of method 1) generated by MicMac.

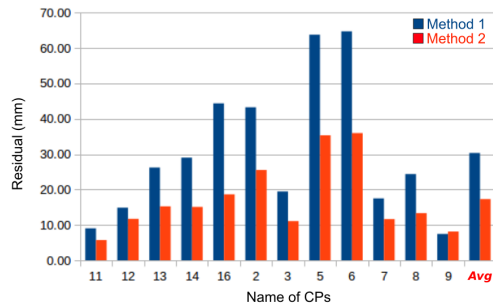


Figure 17. Accuracy check on CPs result- Case 2

CONCLUSIONS AND PERSPECTIVES

In this paper, we present a detailed description of a method that improves the pose estimation precision. The method relies on the result of the classical photogrammetric processing (first iteration) and a supplementary high quality tie point extraction algorithm. The approach proves its efficiency on two real test cases with different scenarios. For a deeper evaluation, the method must be tested on large complex indoor scenario, in which it will probably perform better.

Moreover, the method could be used to accelerate tie point extraction procedure. A first photogrammetric processing with tie points extracted by SIFT at a low resolution can firstly generate an approximate model, then the pose estimation can be refined with the proposed method.

Another possible application is to refine the orientation of satellite images with tie point extraction, which comes with pre-orientation. A quick test of tie point extraction on a crop zone of diachronic satellite image pair compared to SIFT is shown in Figure 18.

APPENDIX

Point	X (mm)	Y (mm)	Z (mm)	D (mm)
1003	1.15	0.57	1.45	1.94
1005	0.17	0.68	0.86	1.11
1007	0.30	0.57	0.57	0.86
1008	0.46	0.87	0.16	0.99
1010	0.05	0.38	0.79	0.88
1013	0.08	0.24	0.02	0.25
1014	1.82	2.15	0.23	2.83
1017	1.08	1.85	0.11	2.14
1018	0.86	1.87	1.85	2.77
1019	0.44	1.93	1.53	2.50
1020	0.23	0.78	0.07	0.82
Avg	0.60	1.08	0.69	1.55
Std	0.56	0.71	0.66	0.91

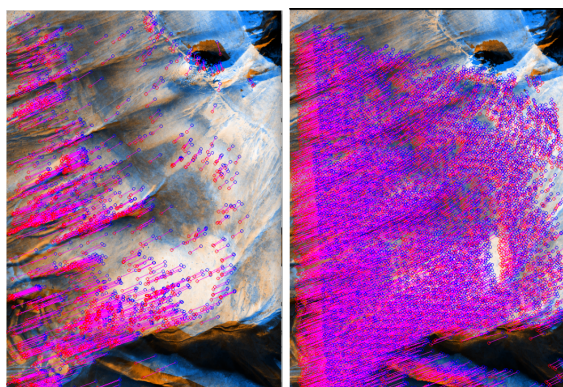
Table 1. Residual of CPs in detail of Case 1-Method 1.

Point	X (mm)	Y (mm)	Z (mm)	D (mm)
1003	0.50	0.46	0.68	0.96
1005	0.17	0.11	0.29	0.35
1007	0.11	0.02	0.42	0.43
1008	0.14	0.07	0.12	0.20
1010	0.34	0.01	0.37	0.50
1013	0.51	0.55	0.06	0.75
1014	0.64	0.61	0.23	0.92
1017	0.08	0.90	0.24	0.93
1018	0.10	0.17	0.62	0.65
1019	0.17	0.58	0.68	0.91
1020	0.07	0.39	0.12	0.42
Avg	0.26	0.35	0.35	0.64
Std	0.21	0.29	0.23	0.27

Table 2. Residual of CPs in detail of Case 1-Method 2

Point	X (mm)	Y (mm)	Z (mm)	D (mm)
11	0.41	2.57	8.64	9.02
12	4.57	1.04	14.09	14.85
13	4.16	1.76	25.83	26.22
14	5.56	2.83	28.33	29.01
16	5.01	0.68	44.08	44.37
2	7.93	4.84	42.26	43.27
3	1.79	4.41	18.85	19.44
5	0.08	0.38	63.78	63.78
6	0.46	2.98	64.65	64.72
7	5.10	2.19	16.59	17.49
8	4.81	0.76	23.88	24.37
9	2.07	7.09	0.92	7.44
Avg	3.50	2.63	29.32	30.33
Std	2.48	2.00	20.45	19.59

Table 3. Residual of CPs in detail of Case 2-Method 1



bridge. Available from: <http://mi.eng.cam.ac.uk/er258/index.html>.

Shragai, Z., Even-Paz, A. and Klein, I., 2011. Automatic tie-point extraction using advanced approaches.

Figure 18. Tie point extraction on diachronic satellite image pair. Left: with SIFT, right: with the proposed method.

Point	X (mm)	Y (mm)	Z (mm)	D (mm)
11	1.96	2.02	4.94	5.69
12	2.77	0.42	11.37	11.71
13	4.24	1.77	14.53	15.24
14	4.06	2.24	14.33	15.07
16	5.05	0.82	17.91	18.63
2	4.88	5.11	24.51	25.51
3	0.61	2.09	10.84	11.05
5	0.62	0.32	35.33	35.33
6	0.09	3.27	35.81	35.96
7	0.65	5.59	10.18	11.63
8	2.64	0.82	13.03	13.32
9	1.46	7.96	0.70	8.12
Avg	2.42	2.70	16.12	17.27
Std	1.79	2.37	10.84	9.95

Table 4. Residual of CPs in detail of Case 2-Method 2

REFERENCES

Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L., 2008. Speeded-up robust features (surf). *Computer vision and image understanding* 110(3), pp. 346–359.

Deseilligny, M. P. and Clery, I., 2011. Apero, an open source bundle adjustment software for automatic calibration and orientation of set of images.

Lingua, A., Marenchino, D. and Nex, F., 2009. Performance analysis of the sift operator for automatic feature extraction and matching in photogrammetric applications. *Sensors* 9(5), pp. 3745–3766.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60(2), pp. 91–110.

Martin, O., Meynard, C., Pierrot-Deseilligny, M., Souchon, J. and Thom, C., 2014. Réalisation dune caméra photogrammétrique ultralégère et de haute résolution.

Morel, J.-M. and Yu, G., 2009. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences* 2(2), pp. 438–469.

Rosten, E., 2006. Fast corner detection. *Engineering Department, Machine Intelligence Laboratory, University of Cam-*