FEATURES OF AERIAL TRIANGULATION BY USING DIFFERENT-TIME IMAGES OF URBAN AREAS

Kobzev A. A.*, Chibunichev A. G.

Moscow State University of Geodesy and Cartography (MIIGAiK), Moscow, Russia – ant50044@yandex.ru

Technical Commission II

KEY WORDS: aerial triangulation**,** neural network, tie points, different-time images, urban areas.

ABSTRACT:

It is well-known that in order to obtain up-to-date spatial information of rapidly developing cities, aerial photography is carried out almost annually. This work is devoted to the joint aerial triangulation of aerial urban area photographs obtained at different times. The main problem of the joint aerial triangulation of different-time images is the process of tie points detection. In this paper it is proposed to search for tie points exclusively on the roofs of building since they are least susceptible to change over time. In order to do this, the roofs of buildings are highlighted on the original aerial photographs via the Unet neural network and then tie points are detected within these areas. The technology made it possible to improve the quality of aerial triangulation: remove photogrammetric gaps in the given block, increase the number of tie points, reduce the processing time by 25% without increasing computing requirements. This approach to the search for tie points made it possible to increase the efficiency of aerial triangulation not only when processing archival and current images together, but also when processing only the results of actual aerial photography.

1. INTRODUCTION

According to the UN report (UN, 2019), there is rapid urbanization with the number of megacities in the World increased from 10 to 43 over the 1990 – 2018 period. Urban development and city management require high-quality geospatial data (SD), which are created mainly as a result of aerial photography or its combination with ground-based survey methods (for example, mobile scanning survey (Zhaojin et al., 2020). Moreover, it is often a problem to perform joint aerial triangulation of space and aerial photos of the same urban area taken in different years (Niu, 2007; Zharova, 2017). According to the research on the topic (Chibunichev and Kobzev, 2021), archival aerial photography can be used for georeferencing of new aerial photographs and additional control. It is known that the main problem of the joint aerial triangulation of different-time images is the process of searching for tie points. 95% of the tie points for different-time images are located on the images of roofs, since they least susceptible to change over time. Therefore, in this paper it is proposed to choose tie points exclusively on the roofs of buildings. The roofs of buildings are detected in the images (special masks are created) with the help of a neural network, and the further search for tie points is carried out within areas covered with these masks. The experimental studies performed on the images of a city have shown the effectiveness of the proposed aerial triangulation technology not only when processing archival and current images together, but also when processing only the results of actual aerial photography.

2. MASKS CREATION

The search for roofs of houses is a special case of solving the problem of image segmentation. Many researchers have used neural networks to interpret the results of aerial photography (Wim van Wegen, 2022; Torres et al., 2020), remote sensing (Condorelli et al., 2021; Carpentier et al., 2021) and to edit three-dimensional models of objects (Andrade et al., 2021).

In our case, the standard architecture of the Unet neural network was applied. It consists of two parts: encoding and decoding. The encoding part of the architecture is a convolutional neural network with sequential convolution and resolution reduction operations. As the result of these operations, the described features of the desired object are determined. The decoding part converts the selected features into the necessary images using evolution and resolution enhancement operations.

The main parameter of this architecture is the number of neurons in the so-called "bottleneck", which connect both parts of the architecture. The "bottleneck" strongly affects the generalizing ability of architecture: if there are not enough neurons, the neural network will not be able to generalize the task; if there are many neurons, the neural network can learn the input data.

Based on the above, it can be seen that the architecture is fully convolutional, allowing you to vary the size of the input images of a the Unet neural network during its operation, since convolution operations do not depend on the size of the submitted images (the number of processed images at a time depends on the amount of video memory). During the neural network training, the size of image fragments is set to 512x512. Figure 1 shows a scheme of the neural network.

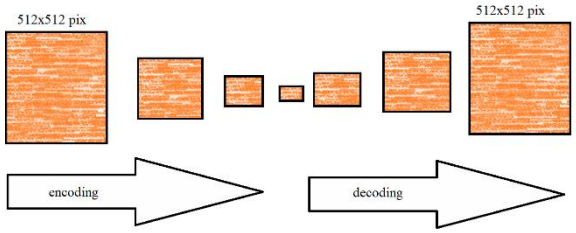

Figure 1. Scheme of encoding and decoding

The main advantage of the chosen architecture is the transfer of calculated features at the encoding stage immediately to the

* Corresponding author

decoding stage. This allows you to speed up the learning process and, as a result, reduce the amount of data needed for training.
Therefore, the contours of buildings were manually selected in three images to create the masks; after that the neural network was trained on this data set and the creation of masks with different binarization (from 0.5 to 0.9) coefficients was launched. Figure 2 shows the comparison of the masks created by the neural network with different coefficients.
Experimentally, it was determined that the binarization coefficient of 0.7 is optimal (on the one hand, it does not give a lot of noise, i.e. extra pixels, and on the other hand, with this coefficient most of the buildings are detected).
There was no requirement for the neural network to accurately determine the contour because in this task it is only necessary to select the areas within which the search for tie points will be performed.

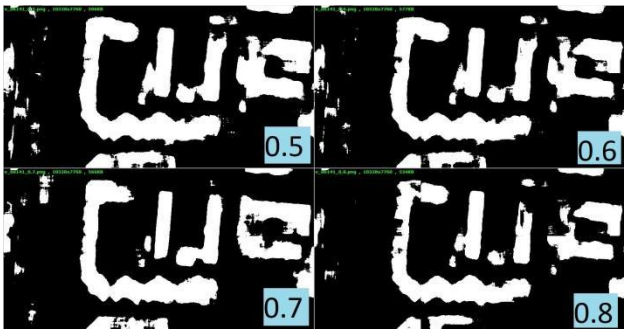

Figure 2. The comparison of the masks with different binarization coefficients. The coefficient is shown in the lower right corner of each mask.

A mask is an image, in which the number of pixels is equal to the number of pixels of the original aerial image for which the mask was made, and pixels take values either 0 = ‘Not a building’, or 255 = 'Building’. Figure 3 shows an example of a mask.

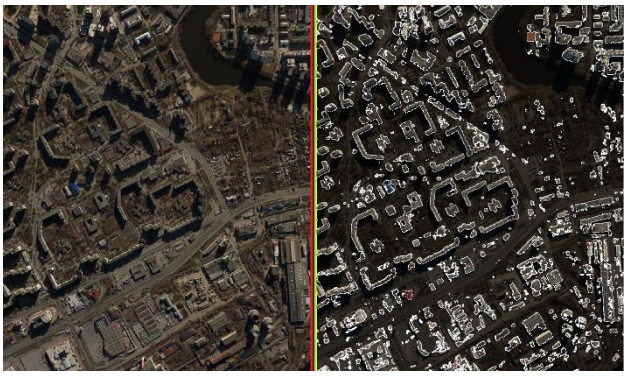

Figure 3. The original aerial photo (left) and the same photo with the superimposed mask (right)

Digital photogrammetric workstations (DPW) allow masks to be used in this form, however, sometimes they need to be inverted.
The neural network training process took about 15 hours on the NVIDIA 2070super graphics video accelerator. After training, the creation time for one mask averaged 3 seconds, that is, 30 minutes for all 607 images participating in the experiment.

3. EXPERIMENTAL STUDY

3.1 Source data

Two sets of different-time aerial photos of the central part of a city were used as initial data for experimental studies. Their characteristics are presented in Table 1.

Characteristic	Archive aerial photography	Actual aerial photography
Year of aerial photography	July 2015	May 2017
Aerial photography system	PhaseOne IXA180	DMC II 250
Height, m	1500	2000
GSD, m	0.1	0.1
Overlap, %	70/40	70/40
Physical pixel size, microns	5.2	5.6
Focal length, mm	80	112
Frame size, pixels	10328 x 7760	14016 x 16768
Frame size on the ground, m	1110 x 820	1690 x 1410
Number of images	436	171
Geodetic reference	GNSS projection centers in WGS84	GNSS+IMU centers+angles in WGS84
Camera calibration	Self-calibration during aerial triangulation	Fixed according to the passport
Ground verification points, pcs.	50	

Table 1. Characteristics of initial aerial photos.

As can be seen from the table, 2 years have passed between aerial photographs. The images captured with different cameras but with the same pixel size on the ground.
Figure 4 shows the experimental area and the distribution of ground check points.

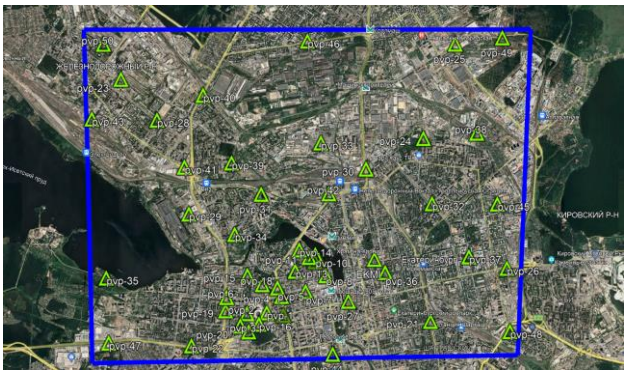

Figure 4. The experimental area and ground check points.

Solid contours (corners of curbs, manhole covers, etc., see example on figure 5) was used as check points. The points were measured using JAVAD Triumph-1M GNSS receivers in fast static mode from two constantly operating reference stations. The processing of differential measurements and the calculation of coordinates in the WGS84 coordinate system was performed in the Waypoint GrafNet 8.7 software package. The average

standard deviation in the plan coordinates was 1.1 cm, in height 1.3 cm.

Figure 5. An example of a check point.

To study the possibility of improving the quality of aerial triangulation when using masks, 12 block adjustments were performed: a separate block of images obtained by the PhaseOne camera, a separate block of images obtained by the DMC camera, and both blocks of images together, with and without masks. The aerial triangulation was performed with various settings (high and medium quality) in Agisoft Metashape Professional.

Evaluation of the results of the search for tie points in images of different years was performed in a special software developed in Python 3.6, and statistical processing parameters were collected and analyzed in MS Excel.

3.2 Aerial triangulations

Table 2 shows the aerial triangulation options. The following settings were set for all blocks: reference and general preselection - enabled, the number of key points per image- 200,000, tie points per stereo pair - 5,000. The quality and mask settings are also shown in Table 2. Masking was performed at the stage of searching for tie points. The "-m" mark additionally indicates the use of masks in the option. Figure 6 shows an example of measured tie points without the use of masks and with them.

№ options	DMC images	PhaseOne images	Masking	Quality
1	+	-	-	High
1-m	+	-	+	High
2	+	-	-	Average
2-m	+	-	+	Average
3	-	+	-	High
3-m	-	+	+	High
4	-	+	-	Average
4-m	-	+	+	Average
5	+	+	-	High
5-m	+	+	+	High
6	+	+	-	Average
6-m	+	+	+	Average

Table 2. Aerial triangulation options.

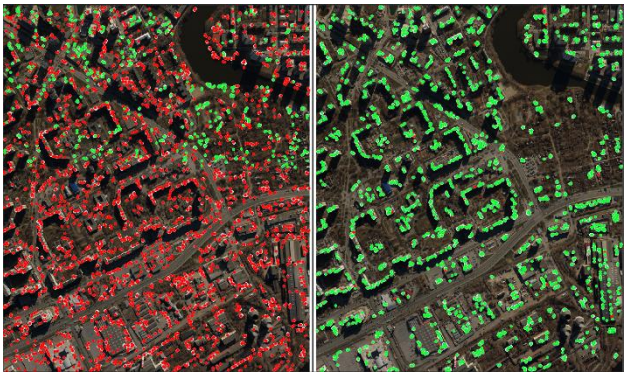

Figure 6. Tie points on the image without masks (left) and with masks (right). The green dots are final points, the red ones are filtered and removed.

As can be seen in Figure 6, besides the fact that the tie points when using masks (on the right) were found only on the roofs of buildings, also most of the points are not filtered (not rejected), unlike the case when masks are not used (on the left).

Figure 7-a and 7-b show the tie points detected at the same area with and without masking original images.

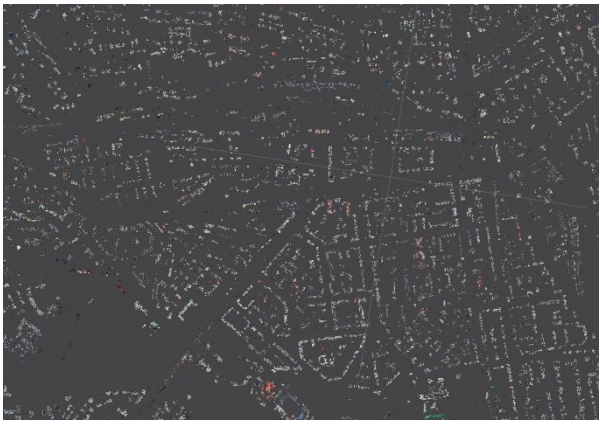

Figure 7-a. Tie points found using masks

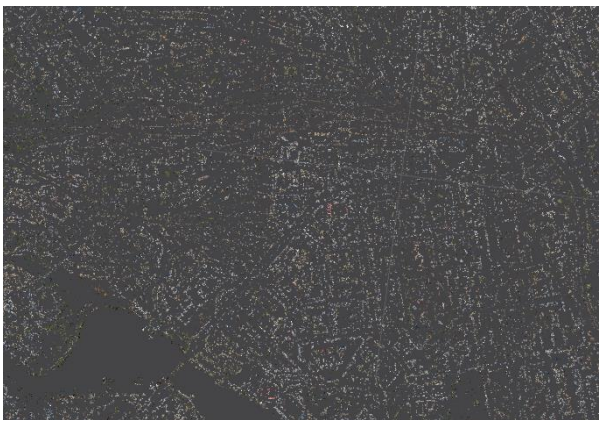

Figure 7-b. Tie points found without using masks

3.3 Methodology of evaluating results and its criteria

Aerial triangulation was evaluated according to the following criteria:

- Percentage of oriented images;
- Total number of tie points;
- Number of tie points measured in images of different years;

- Average reprojection error;
- Processing time;
- Expended computer memory resources;
- Errors at ground check points.

The pixel coordinates of the ground check points were measured in all the images in advance and then they were imported into the project of each aerial triangulation option. This method guaranteed the exclusion of errors due to different observations of the same check point in different options.

3.4 The software for analyzing the results of the tie points search

In order to visually control the location of tie points, especially in images taken at different times and with different cameras, a special software was developed. It gets tie points in Bundler format and lists of archived and actual images, as well as the path to the source images as input data. The software goes on from a point to a point in the Bundler file and select the names of the images on which a tie point was measured from the list using indices. Then it checks whether the name is in both archive and actual lists, if so, the point is to be remembered. At the output, the program generates a text file with statistics by points (in how many images each tie point is measured) and by images (how many tie points are in the image). Figure 8 shows the program interface.

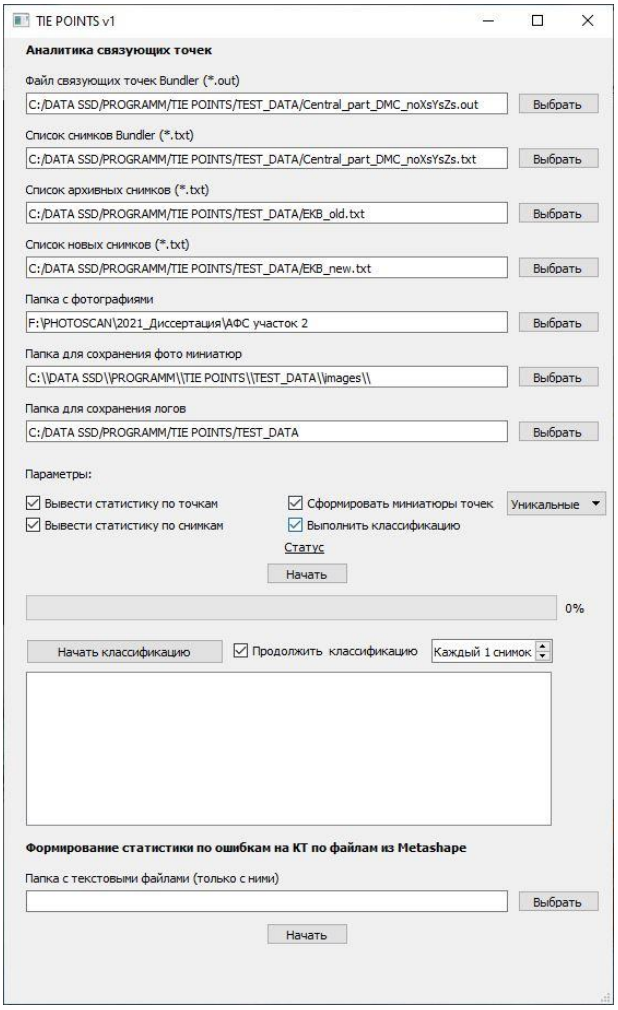

Figure 8. Tie points analyzer program interface.

After selecting the points, the software can also cut out image frames (see Figure 9) with tie points in order to be able to assess which terrain object a particular point belongs to.

Figure 9. An example of an image frame with a tie point on the roof of a building generated by the software.

Generating such image frames allows to compare tie points (see figure 10).

Figure 10. An example of a tie point in different years images (a – PhaseOne IXA180, b – DMC II 250).

Then the frames with the tie points can be manually classified into classes (roofs, urban buildings, forests or fields, other). To do this, the software shows a specific frame and wait for pressing the digit of the class number on the keyboard. This process is necessary to check the correctness of the mask creation to make sure that the tie points are really found only on the roofs.

The experiment has shown that it takes 10 minutes for the program to process 1 million tie points and select approximately 30,000 points detected in different-time images out of them; it takes 4 hours to generate all image frames with the tie points.

3.5 Evaluation of aerial triangulation results

Tables 3-5 show the results of aerial triangulation for each variant according to different criteria.

№	The total number of tie points	The number of tie points between years	The average reprojection error, pixels
1	210,766	-	0.27
1-m	281,275	-	0.31
2	229,481	-	0.46
2-m	308,341	-	0.73
3	757,683	-	0.34
3-m	772,021	-	0.43
4	727,776	-	0.55
4-m	658,319	-	0.90
5	1,115,776	13,352	0.29
5-m	1,148,815	38,921	0.32
6	947,633	14,943	0.40
6-m	1,046,514	35,216	0.59

Table 3. The results of the search for tie points.

№	RMS X, cm	RMS Y, cm	RMS Z, cm	RMS plan, cm
1	54	8	23	55
1-m	28	8	23	29
2	8	6	21	10
2-m	9	6	20	11
3	8	8	23	11
3-m	7	8	21	11
4	7	7	23	10
4-m	7	7	19	10
5	8	7	23	11
5-m	13	7	26	15
6	8	8	23	11
6-m	9	8	24	12

Table 4. The accuracy verification for 50 ground check points.

№	Time, s	Time, H:MM	RAM, GB
1	4260	1:11	8.43
1-m	4080	1:08	9.23
2	4380	1:13	4.41
2-m	3660	1:01	4.6
3	1827	0:30	3.15
3-m	1032	0:17	3.21
4	937	0:15	1.54
4-m	538	0:09	1.45
5	8280	2:18	9.51
5-m	7560	2:06	9.81
6	6600	1:50	4.24
6-m	4440	1:14	4.56

Table 5. The resources expended.

As can be seen from Table 4, the first variant (DMC, high quality) of aerial triangulation showed low accuracy. This is due to the fact that the exterior orientation elements were determined only for 139 aerial images out of 171. On average, 87% of the tie points per image were filtered out of 32 not-oriented aerial photographs. The use of masks (the second option) made it possible to determine the exterior orientation elements of all images in the block. At the same time, only 6% of the tie points were rejected.

For next diagrams percentage was calculated as reference between the characteristic number calculated in option with a mask and the number without a mask.

4. RESULTS

As the result of the experiment, the quality of block adjustment was analyzed according to a variety of criteria: the total number of tie points and the number of tie points between aerial images of different years, the reprojection error, the processing time and the amount of RAM used, the error at ground check points, etc.

Having analyzed the results of the experiment, the following conclusions have been made:

1. In general, the use of masks does not affect the accuracy of aerial triangulation. In some cases, when photogrammetric discontinuities appear as the result of aerial triangulation (for some images of the block, the exterior orientation elements are not determined), the use of masks solves the problem (i.e., exterior orientation elements are determined for all images) resulting in the increased accuracy of aerial triangulation.
2. The total number of tie points when using masks increases by an average of 12%, a maximum of 34%. The number of tie points between aerial images of different years increases 2.6 times when using masks. Such an increase is associated with the initial search for reliable points on the roofs of buildings which were less effected by external conditions, that is especially noticeable after several years.
3. The time for performing aerial triangulation using masks is reduced by an average of 25% (from 5% to 44%) because buildings in urban areas occupy about a quarter of aerial image.
4. The use of masks does not affect the requirements for computer hardware (on average, this increases the use of RAM by 3%).

6. CONCLUSION

Based on the above, it can be recommended to use masking of aerial images at the stage of searching for tie points when processing aerial photos of urban areas, especially when processing different-time images. This allows you to improve the quality of aerial triangulation, avoid photogrammetric discontinuities and reduce processing time without the need of additional resources.

ACKNOWLEDGEMENTS

Authors wish to special thank LLC “Ural-Siberian Geospatial Company” for providing the source data for the research and measuring ground check points.

REFERENCES

Andrade, A. C., Alixandrini Jr., M. J., Carvalho, F. P. S., and Fernandes, V. O.: PERFORMANCE OF THE SUPPORT VECTOR MACHINE AND ARTIFICIAL NEURAL NETWORK CLASSIFIERS FOR ROADS IDENTIFICATION. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2021, 9–14, https://doi.org/10.5194/isprs-archives-XLIII-B3-2021-9-2021, 2021.

Carpentier, B., Masse, A., Lavergne, E., and Sannier, C.: BENCHMARKING OF CONVOLUTIONAL NEURAL NETWORK APPROACHES FOR VEGETATION LAND COVER MAPPING. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2021, 915–922, https://doi.org/10.5194/isprs-archives-XLIII-B2-2021-915-2021, 2021.

Chibunichev A.G., Kobzev A.A., 2021: Possibility of joint photogrammetric processing of different-time aerial photos. *Izvestia vuzov «Geodesy and Aerophotosurveying»*, 65 (3), 292–301. [In Russian]. DOI:10.30533/0536-101X-2021-65-3-292-301

Condorelli, F., Rinaudo, F., Salvadore, F., and Tagliaventi, S.: A COMPARISON BETWEEN 3D RECONSTRUCTION USING NERF NEURAL NETWORKS AND MVS ALGORITHMS ON CULTURAL HERITAGE IMAGES. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2021, 565–570, https://doi.org/10.5194/isprs-archives-XLIII-B2-2021-565-2021, 2021.

Torres, D. L., Feitosa, R. Q., La Rosa, L. E. C., Happ, P. N., Marcato Junior, J., Gonçalves, W. N., Martins, J., and Liesenberg, V., 2020: Semantic segmentation of endangered tree species in brazilian savanna using Unet variants *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3/W12-2020, 355–360, https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-355-2020

United Nations, Department of Economic and Social Affairs, Population Division (2019). World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). New York: United Nations

Wim van Wegen, 2022. 5 Questions to… Jolle Jelle de Vries, Geomaat. *GIM International.* https://www.gim-international.com/content/article/ai-definitely-is-not-just-a-hype

Xutong Niu, Feng Zhou, Kaichang Di, and Rongxing Li, 2007: 3D geopositioning accuracy analysis based on integration of QuickBird and IKONOS imagery. *Photogrammetric Engineering and Remote Sensing.*

Zhaojin L., Bo W., Yuan L., 2020: Integration of aerial, MMS, and backpack images for seamless 3D mapping in urban areas. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020.

Zharova N., 2017: Criteria for forming stereo "fortuitous" satellite image pairs. Accuracy estimation of products created using stereo "fortuitous" satellite image. *Izvestia vuzov. Geodesy and aerophotography*, 6, 59-68.