Aerial Triangulation Using Different Time Images of Urban Areas Obtained from Unmanned Aerial Systems

Anton A. Kobzev[1], Alexandr G. Chibunichev[2]

[1]JSC Ural-Siberian Geo-Information Company, Yekaterinburg, Russia – ant50044@yandex.ru
[2]Moscow State University of Geodesy and Cartography (MIIGAiK), Moscow, Russia – agchib@mail.ru

KEY WORDS: Neural network, spatial data, tie points, key points, aerial photography, UAV, masking, urban areas.

ABSTRACT:

In case to ensure the development of urban areas, it is necessary to create and promptly update spatial data, which are primarily created using aerial photography. With the development of unmanned aerial systems, the survey of cities is mainly carried out with their help. However, due to the large amount of interference or failures in the operation of GNSS equipment installed on the UAV, the aerial photographs miss the coordinates of the projection centers or their quality does not allow to accurately process the photos for further creation of spatial data. It leads to perform extra field work, which affect time and cost of production.
To solve this problem, we propose the technology of using archival aerial survey data as a source for geodetic reference in the alignment of the new aerial images for same areas, which allows you to reduce the time and cost of work. Therefore, the purpose of this work is to investigate the possibility of the aerial triangulation using different time images of urban areas obtained from UAVs.
The research shows the results of experimental study on UAVs survey of two urban areas and includes flights made in 2018-2024 years with GSD 5-7 cm. In order to find tie points, we limit the searching area for key points using masks generated with the neural network YoloV8_p6. The accuracy of the resulted alignment was controlled using ground-based check points.
The performed research allowed us to confirm the validity of the proposed technology for using archival aerial photos to geo-reference new ones not only for aerial survey from manned aircrafts but also for photos from UAVs. The technology of aerial triangulation using different time images of urban areas can be recommended for aerial survey from manned vehicles for cities with population of millions of people, as well as aerial survey from unmanned vehicles when updating less populated city's spatial data.

1. INTRODUCTION

In case to ensure the development of urban areas, it is necessary to create and promptly update spatial data. Today, these data are primarily created using aerial photography (Khawte et al., 2022; Wahbeh et al.m 2022). With the development of unmanned aerial systems, the survey of cities is mainly carried out with their help (the only exception is the survey of megacities with skyscrapers or an area of more than 1000 sq. km) (Latha et al., 2019).
However, due to the large amount of interference or failures in the operation of GNSS equipment installed on the UAV, the aerial photographs miss the coordinates of the projection centers or their quality does not allow to accurately process the photos for further creation of spatial data. In such cases, it is a common practice to carry out aerial survey once again or perform extra field work - additionally measure the coordinates of the control points using GNSS. To solve this problem, we propose the technology of using archival aerial photography data as a source for geodetic reference in the alignment of the new aerial photos for same areas, which allows you to reduce the time and cost of work. Similar ideas have been suggested before but mainly concerned the joint use of space imagery (Niu et al., 2007; Zharova, 2017) and integration with ground-based survey methods (Zhaojin et al., 2020).
The research has been conducted with the use of the technology to geo-reference new aerial photography from manned aviation vehicles (Chibunichev and Kobzev, 2021; 2022). However, currently cities are surveyed with UAVs, which have their own peculiarities, for example, tilt angles can be up to 20 degrees, which naturally makes it difficult to find tie points especially on different time photos. Therefore, the purpose of this work is to investigate the possibility of the aerial triangulation using different time images of urban areas obtained from UAVs.

2. EXPERIMENTAL STUDY

The experimental research consisted in performing a joint aerial triangulation based on different time images obtained from an unmanned aircraft for two settlements. The first trial survey of a settlement (about 10 thousand people population, area of 12 sq. km) was performed using an unmanned aircraft Geoscan-201 with a Sony RX1 camera (24MP, GSD 7 cm) in 2018 and 2019. The aerial triangulation was carried out using Agisoft Metashape software. The GNSS measurements were processed in the Waypoint GrafNav 8.7 software. The 2018 images were used as a reference for the triangulation of the 2019 survey. The accuracy of the aerial triangulation was assessed using 8 check points measured in images from 2019 only, which corresponded to the accuracy of the alignment of the archival aerial photography of 2018.
It is important to mention that most of the tie points between different time images were found mainly on the roofs of buildings (figure 1). We later used this fact to find tie points by extracting the roofs of houses in the images and searching for tie points only in the selected areas. This experiment showed a real possibility of using archival images as a reference for joint aerial triangulation based on images taken from UAVs.
These results served as a basis for more detailed research, which we describe in a detail in this article.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W5-2024-87-2024 | © Author(s) 2024. CC BY 4.0 License.

Figure 1. Sample of Tie Points Measurements

A city with a population of about 75 thousand people, an area of 243 12 sq. km, was used as the object of the research. Aerial survey was performed in 2022 and 2024 using the Geoscan-201 UAV and the Sony RX1RM2 camera (42MP, GSD 5 cm), the Topcon B111 GNSS receiver. The number of images in 2022 was 7625, in 2024 – 8643, the reference was carried out with the coordinates of the projection centers for aerial photographs with the evaluation using ground check points. In total, 70 check points were measured by ground-based GNSS survey equipment. The check points were selected in such a way that they appear on the images of 2022 and 2024. For example, curbs and manholes were used. All points were photographed from several angles to ensure unambiguous identification in the images. Multi-frequency multi-system satellite receivers EFT RS2 in RTK mode were used for measurements. To ensure the uniformity of measurements, they were performed from the same constantly operated reference station.

An example of a measured ground point is shown in figure 2.

The characteristics of aerial survey are presented in the table 1.

When performing aerial survey in 2024, there were failures in the operation of GNSS receivers over the city center, so several routes were not provided with high-precision projection centers.

GNSS trajectory processing and calculation of the coordinates of the projection centers of aerial images were performed using a differential method relative to the reference stations installed at the UAV takeoff site. The calculations were carried out in the Waypoint GrafNav software.

To control the accuracy of the geodetic reference, 70 control points were measured by ground-based GNSS survey equipment. The check points were selected in such a way that they appear on the images of 2022 and 2024. For example, curbs and manholes were used. All points were photographed from several angles to ensure unambiguous identification in the images. Multi-frequency multi-system satellite receivers EFT RS2 in RTK mode were used for measurements. To ensure the uniformity of measurements, they were performed from the same constantly operated reference station. As a result, a catalog of points was created, containing their names and coordinates.

Characteristic	**Archive aerial photography**	**New aerial photography**
Year of aerial photography	2022	2024
Aerial photography system	Sony RX1RM2 42MP	Sony RX1RM2 42MP
Height, m	330-370	310-389
GSD, cm	5	5
Overlap, %	80/70	80/70
Physical pixel size, microns	4.5	4.5
Focal length, mm	35	35
Frame size, pixels	5304 x 7952	5304 x 7952
Frame size on the ground, m	265 x 398	265 x 398
Number of images	7625	8643
Geodetic reference	GNSS projection centers in WGS84	GNSS projection centers in WGS84
Camera calibration	Self-calibration during aerial triangulation	Self-calibration during aerial triangulation
Ground verification points, pcs.	Total 70 (22 used for the research)	

Table 1. Characteristics of initial aerial photos.

Figure 2. Sample of a check point measurements

In order to assess the accuracy of aerial triangulation of 2022 and 2024 surveys separately, all the points were duplicated and suffixes were added to their names: "im2024" and "im2022", meaning that this particular point was measured in the images of 2024 or 2022. This made it possible to perform a separate assessment of the external orientation of the 2022 and 2024 images using the same points.

2.1 Masks Creation

As it was performed in the previous research on manned aerial photography (Chibunichev and Kobzev, 2022), to limit the area for key points to the roofs of buildings, we used masks, created using the Deeplabv3+ neural network (Torres et al., 2020), to improve the search for tie points. A mask is created for each aerial image and represents a black-and-white image equal to the original image in the number of pixels. The white pixels on the mask mean that this pixel belongs to the object of interest (the roof of the building) and should be used when searching for tie points, and the black one does not. An example of the mask is shown in figure 3.

Figure 3. Sample of a mask

The YoloV8 neural network was also trained to improve the quality of masks. The distinct advantage of this neural network is that it can highlight clearer outlines of objects without rounding them off. However, its main disadvantage turned out to be poor work with large objects: if an object in the image takes up most of the image, the neural network does not recognize it well. This problem arises when processing small images from UAVs on the territory of a city where there is a large number of apartment buildings.

The solve this problem we chose the "YoloV8_p6" neural network architecture and increased the maximum size of the selecting object.

To mark up data for neural network training, the CVAT service was deployed. It is a free web-based open-source image and video annotation tool used to label data for computer vision algorithms.

CVAT supports the main tasks of supervised machine learning:

- object detection;
- image classification;
- image segmentation.

The tool has many functions, including:

- interpolation of shapes between keyframes;
- semi-automatic annotation using deep learning models;
- shortcuts for most important actions;
- a dashboard with a list of projects and annotation tasks;
- LDAP and basic access authentication.

CVAT is distributed under the MIT license, and its source code is available on GitHub (figure 4).

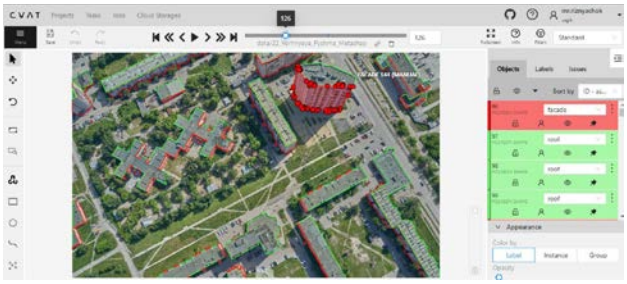

Figure 4. Markup using CVAT

The 30 images marked up in CVAT were used to train both neural networks.

A special server was devoted to the neural network trainings and generating masks. The DeepLab neural network training took 4 hours on one NVIDIA RTX 2070 Super video accelerator, the Yolo training took 17 hours on four NVIDIA Tesla T4 video accelerators. The peculiarity of the Yolo training was that the extraction was performed not only for roofs, but also for facades. For this research, masks for facades and roofs were combined into one, but in the future, their separation can be used to improve technology and solve other problems. Figure 5 shows the result of extracting roofs (in red) and facades (in blue), as well as an estimated probability that the selected fragment belongs to the specified class.

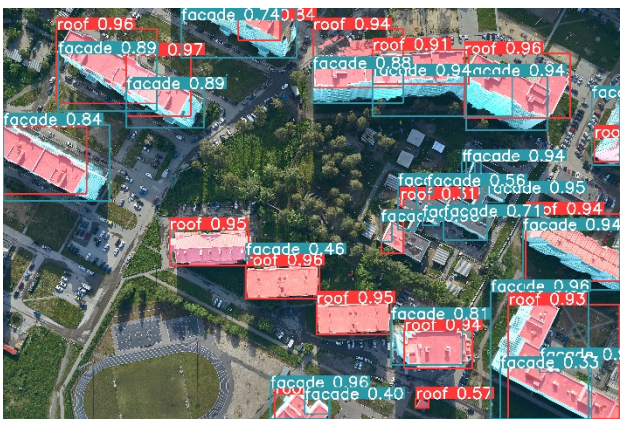

Figure 5. Sample of a mask for roofs and facades with classes

After training neural networks, masks were generated for all 4,000 images using both neural networks on the same NVIDIA Tesla T4 video accelerator. The speed of creating masks using DeepLab with the optimization of the process of their generation is 3 seconds per mask, the Yolo model is 9 seconds per mask without optimizations.

The masks for the same image obtained from the DeepLab model (figure 6) and the Yolo model (figure 7) are presented below.

From what we can see in the figures, we come to a conclusion that despite the fact that the DeepLab training and inference are faster and it consumes fewer resources, the Yolo model gives a much better result in extracting roofs. This is the reason why in further processing it was decided to use masks created by the Yolo model.

Figure 6. Sample of a mask generated using DeepLab.

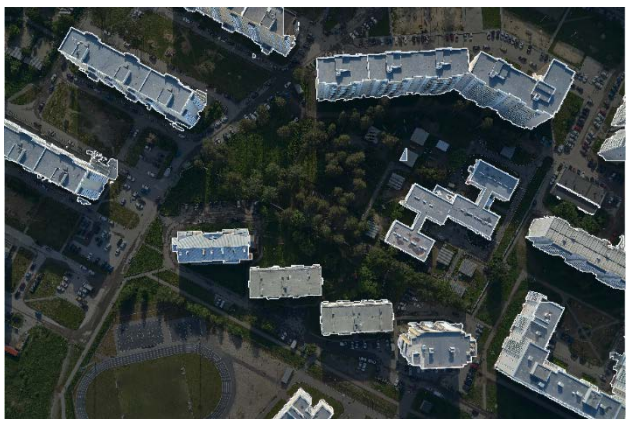

Figure 7. Sample of a mask generated using Yolo.

2.2 Aerial Triangulation and Its Options

Aerial triangulation, including the search for key points in the images, their matching and alignment, was performed in Agisoft Metashape 2.0.3 software. The processing imitated a situation where the images of 2022 had a good geodetic reference, but the images of 2024 did not. Masks were used to limit the search area for key points. The same chunk of images were processed without masks to assess their impact. The aerial surveys of 2022 and 2024 were processed with the projection centers of aerial photographs as a control data.
For aerial triangulation processing we used the following settings:

- The number of key points per image is 40,000;
- The number of tie points per stereo pair is 4,000;
- Preliminary selection by geo-referencing is enabled;
- Masking is enabled for key points.

Table 2 shows the aerial triangulation options.
The "-m” mark in the option name indicates that masks were used in this case.

# Options	2022 Images	2022 Reference	2024 Images	2024 Reference	Masking
1	+	+	+	+	-
2	+	+	+	-	-
3-m	+	+	+	-	+

Table 2. Aerial triangulation options.

Thus, the option #1 is the joint processing of aerial photographs with projection centers for both 2022 and 2024 sets of images. This option was used to evaluate the mutual orientation of the sets of the images and the overall accuracy of the alignment, which can be achieved for archival and new aerial survey.
The option #2 is the referencing the aerial images of 2024 to the images of 2022. The aerial survey in 2022 had the coordinates of the projection centers of aerial photographs, and the 2024 survey did not.
The option #3-m is the same as the option #2, but with the use of masking.
To carry out the experiment, only the images of the city central part that mainly has buildings and faces major changes (due to urban development) were selected from the entire set of the images. There were no changes in the rest of the city, so the pictures of this part of the city were not used. 1,115 images from 2024 and 795 images from 2022 were taken for processing. At the same time, 22 geodetic check points were used to assess the accuracy of the aerial triangulation.

2.3 Analysis of Tie Points

The analysis of tie points was performed in a specially developed program that loops through the entire array of tie points, selects some that are measured on different time images, provides statistical information, and generates a report on each tie point and image. The program shows how many images from different years contain a tie point and how many such tie points are located in each image. This allowed us to assess the quantity of tie points in the images from different years.
The program is written in the Python 3.11 programming language. The time of processing and extracting the tie points measured in images of different years from the entire set of points depends on the total number of points and images. The analysis of 1.5 million connecting points takes approximately 4.5 minutes. The user interface of the program is shown in figure 8.

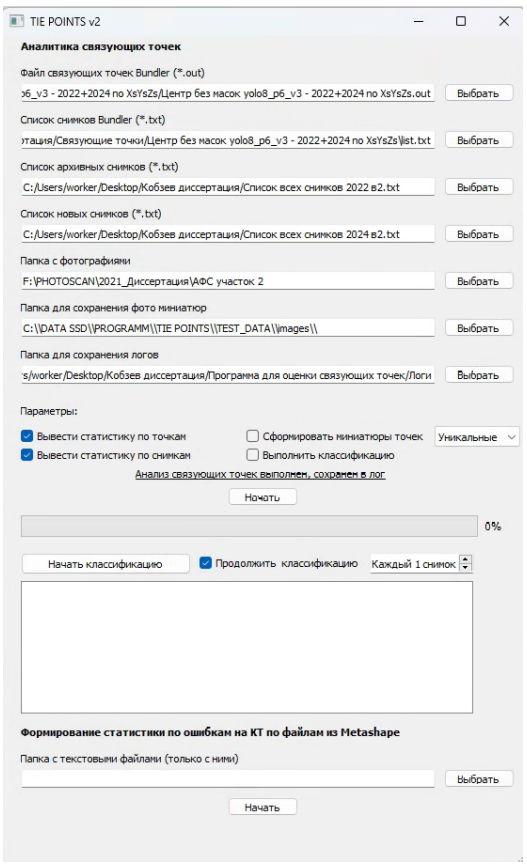

Figure 8. A program for evaluating connecting points.

The results of aerial triangulation by the total number of tie points, the number of points measured in images from different years, as well as an example of a point cloud with and without masks are presented below.

# Options	Total number of tie points	The number of tie points between the years	Masking
1	1,401,484	3,102	-
2	1,383,574	2,457	-
3-m	1,424,828	16,083	+

Table 3. Tie points statistics.

As we can see from table 3, the use of masks allows you to increase the number of tie points measured on both archival and new images **by 7 times**, which contributes to a better geodesic referencing of a new aerial survey to an archival one.

Figure 9. Sample of tie points clouds without (up) and with (down) masking.

Figure 9 clearly shows that the use of masks allows you to find tie points on the roofs of buildings.
The total number of tie points remains approximately at the same level, despite the limitation of the search areas, which guarantees a good connection between the images of one year.
Figure 10 below shows the distribution of the density of tie points between aerial images of different years (the number indicates the number of such tie points per image). The maximum number and the minimum number of the tie points is shown in red and in blue, respectively. The black dots show the coordinates of the projection centers of the 2022 aerial photographs.

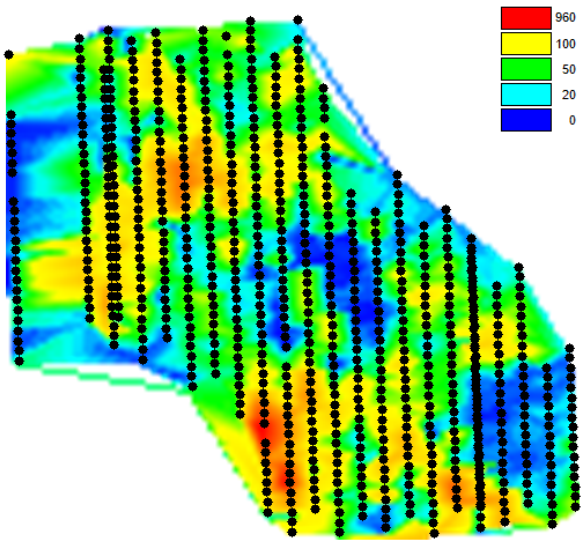

Figure 10. Tie points distribution over the territory.

As can be seen in figure 10, almost the entire area of joint processing is provided with at least 50 tie points per image (green, yellow and red areas). The blue areas (where there are no tie points) correspond to the absence of urban constructions and are also surrounded by images with a sufficient number of tie points.

2.4 Analysis of the Accuracy of Aerial Triangulation by Check Points

As indicated in section 2.3, the geodetic check points in the catalog have been duplicated and suffixes have been added to them, indicating which image of which year a particular point should be measured. Thus, the tie points with the same coordinates but different names appeared in the catalog. For example:
Point_1_im2024, x1, y1, z1
Point_1_im2022, x1, y1, z1
Point_2_im2024, x2, y2, z2
Point_2_im2022, x2, y2, z2
…
The points with the suffix “im2022” were measured from the images of 2022 and allowed us to assess the accuracy of the alignment of the archival survey.
The tie points with the suffix “im2024” were measured only from the images of 2024, allowing us to control how the new aerial survey was aligned to the archival one.
The check points were measured across all the images once, and then exported and imported between the sets to ensure uniformity. The xml exchange format was used for this purpose.
The location of the check points is shown in figure 11 below.

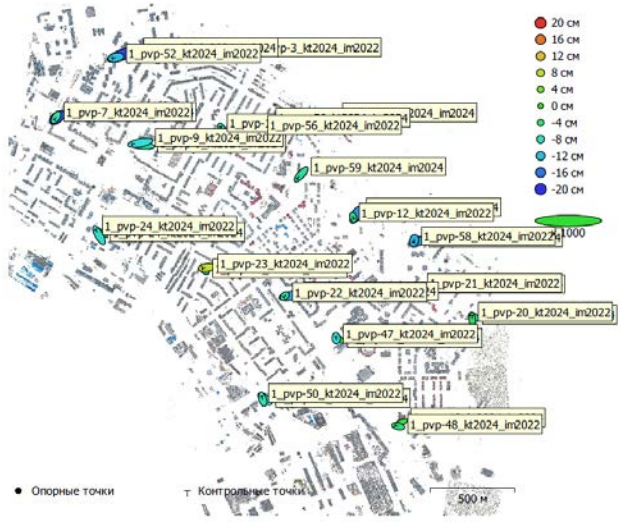

Figure 11. Check points distribution.

Table 4 illustrates the root mean square errors in plan (RMS_S) and altitude (RMS_Z), calculated from the differences of the geodetic coordinates of the check points and obtained as a result of performing aerial triangulation in various options, without masks and with masks according to the proposed technology:

# Options	RMS_S, cm	RMS_Z, cm	RMS_S, cm	RMS_Z, cm
	2022		2024	
1	5	8	7	12
2	5	10	7	46
3-m	5	9	6	10

Table 4. Errors on check points.

As can be seen from table 4, the accuracy of aerial triangulation with reference of the new aerial survey to the archival one, using masking to search for tie points, corresponds to the accuracy of aerial triangulation from archival images using projection centers as control data. Without the use of masks, the altitude error increases significantly, since the number of tie points measured simultaneously on new and archival images decreases significantly (table 3), which leads to worse photogrammetric alignment. Therefore, it is recommended to mask the images.

3. RESULTS

As a result of the work performed, aerial photographs of 2024 without coordinates of the projection centers obtained using an UAV were referred to archival aerial photography of 2022 on the territory of the city. To control the results, the tie points and errors at the check points measured by the ground-based GNSS survey equipment were analyzed.

To ensure a high-quality search for tie points between images from different years, the search area was limited using masks specially created using the YoloV8 neural network. Masks allowed us to direct the algorithm for searching for tie points only to objects that do not change over time – buildings, in particular, their roofs.

This made it possible to significantly increase the number of tie points measured simultaneously on archival and new images and to reference the new aerial survey to the old one.

Thus, we highlight the following research results:

- It is recommended to use the YoloV8_p6 neural network to create masks;
- The use of masks in searching for tie points between images of different years in urban areas allows you to significantly increase the number of tie points (in our case, from 2.5 thousand to 16 thousand points).
- The use of the proposed technology and the masking of the key points search area allows you to reference a new aerial survey to an archival one with an accuracy corresponding to the alignment of the source data.

4. CONCLUSION

The performed research allowed us to confirm the validity of the proposed technology for using archival aerial photos to geo-reference new ones not only for aerial photography from manned aircrafts but also for surveys from UAVs.

ACKNOWLEDGEMENTS

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

REFERENCES

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

Kobzev, A. A. and Chibunichev, A. G.: Features of aerial triangulation by using different-time images of urban areas, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B2-2022, 71–76, https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-71-2022, 2022.

Khawte, S. S., Koeva, M. N., Gevaert, C. M., Oude Elberink, S., and Pedro, A. A.: DIGITAL TWIN CREATION FOR SLUMS IN BRAZIL BASED ON UAV DATA, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-4/W4-2022, 75–81, https://doi.org/10.5194/isprs-archives-XLVIII-4-W4-2022-75-2022, 2022.

Preethi Latha, T., Naga Sundari, K., Cherukuri, S., and Prasad, M. V. V. S. V.: REMOTE SENSING UAV/DRONE TECHNOLOGY AS A TOOL FOR URBAN DEVELOPMENT MEASURES IN APCRDA, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 525–529, https://doi.org/10.5194/isprs-archives-XLII-2-W13-525-2019, 2019.

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 deeplabv3+ 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

Tsay, J.-R., & Lee, M.-S., 2012: SIFT for dense point cloud matching and aero triangulation. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B3. doi.org/10.5194/isprsarchives-XXXIX-B3-69-2012

Wahbeh, W., Müller, G., Ammann, M., and Nebiker, S.: Automatic image-based 3D reconstruction strategies for high-fidelity urban models – comparison and fusion of UAV and mobile mapping imagery for urban design studies, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B2-2022, 461–468, https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-461-2022, 2022.

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. doi.org/10.5194/isprs-archives-XLIII-B2-2020-443-2020

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.