Improving accuracy of photogrammetry method using image segmentation by YOLO neural networks*

Serhiy Balovsyak[1,†], Nazar Hrynyk[1,*,†], Zhengbing Hu[2,†], Khrystyna Odaiska[1,†] and Ivanna Yanchuk[1,†]

[1] *Yuriy Fedkovych Chernivtsi National University, Kotsiubynsky 2, 58012, Chernivtsi, Ukraine*
[2] *School of Computer Science, Hubei University of Technology, Wuhan, China*

Abstract
A review of modern methods and tools for building three-dimensional (3D) models of physical objects has been carried out. An information technology for building 3D models of objects using the photogrammetry method has been proposed, which provides an increase in the accuracy of their construction due to automatic masking of images. The hardware of the technology consists of video cameras, computers and network equipment. The software of the technology consists of the developed program "YOLO_mask_24" for forming masked images, the 3DF Zephyr program for building 3D models using the photogrammetry method. The program "YOLO_mask_24" was developed in Python on the Google Colab cloud platform for segmentation, scaling, dilation, filtering and masking of object images. Image segmentation was performed by a convolutional artificial neural network YOLOv8. 3D models of an object (bench) were constructed using the 3DF Zephyr program based on images without masking. The accuracy of most of the object model is satisfactory, but defects appeared in part of the model. It was found that constructing 3D models using photogrammetry only based on masked images leads to significant model defects.
The novelty of the work is the simultaneous use of images of objects without masking and with masking to build their 3D models using the photogrammetry method. To verify this approach, three-dimensional models of the object (bench) were constructed using the 3DF Zephyr program based on images without masking and with masking. The use of the proposed information technology ensured high accuracy for all parts of the 3D model of the object. The resulting 3D models of objects are used, in particular, in three-dimensional computer graphics, virtual and augmented reality systems.

Keywords
Information technology, 3D model, photogrammetry, image segmentation, image masking, artificial neural networks, software, Python, cloud platform

1. Introduction

The purpose of the photogrammetry method is to build three-dimensional (3D) models of physical objects based on a series of their digital photographs obtained from different angles [1-3]. The photogrammetry method is used to build, in particular, 3D models of geographical and natural objects (terrain relief, landscape, etc.), architectural objects (buildings, structures, etc.), industrial and technical objects (mechanisms, devices, etc.), cultural and artistic objects (sculptures, etc.), biological and medical objects. Such 3D models make it possible to automate and increase the efficiency of solving many applied problems, therefore photogrammetry is widely used in architecture, archaeology, construction, video games, education, medicine, virtual reality and other

Intelitsis'25: The 6th International Workshop on Intelligent Information Technologies & Systems of Information Security, April 04, 2025, Khmelnytskyi, Ukraine
* Corresponding author.
† These authors contributed equally.
s.balovsyak@chnu.edu.ua (S. Balovsyak); hrynyk.nazar@chnu.edu.ua (N. Hrynyk); drzbhu@gmail.com (Zh. Hu); k.odaiska@chnu.edu.ua (Kh. Odaiska); i.yanchuk@chnu.edu.ua (I. Yanchuk)
0000-0002-3253-9006 (S. Balovsyak); 0009-0001-6680-7099 (N. Hrynyk); 0000-0002-6140-3351 (Zh. Hu); 0000-0002-3167-1195 (Kh. Odaiska); 0000-0002-1456-0063 (I. Yanchuk)

industries [4]. The construction of three-dimensional models of objects is performed by specialized programs, for example, 3DF Zephyr [2].

In addition to the photogrammetry method, LIDAR (Light Identification, Detection and Ranging) laser rangefinders are used to build 3D models, the principle of which is to calculate the distance to objects due to delays in the propagation of optical radiation [2]. The use of LIDAR provides high accuracy of 3D models, but requires special hardware and software. Therefore, the photogrammetry method is preferable for solving many problems of building models, the initial images for which can be obtained using video cameras.

A significant problem in the application of the photogrammetry method is that photographs, in addition to the objects under study, often contain images of extraneous objects (background). Such objects in the images can lead to defects in the construction of three-dimensional models. The impact of extraneous objects on the reduction of the quality of 3D models is especially significant in the case of a non-uniform background and in the presence of transparent or specularly reflective surfaces.

One of the ways to increase the accuracy of constructing three-dimensional models using the photogrammetry method is image masking. The essence of masking is to select and further process only certain areas of the image (mask) that belong to the object under study. In this case, other areas of the image outside the mask are usually painted with a uniform color. Thus, masking removes excess information from the original images that can cause artifacts.

In most photogrammetry programs, image masking is performed manually, which is a rather laborious and subjective process [2]. Image masking can be automated using segmentation methods that separate the studied objects as separate segments. Currently, the photogrammetry method is actively developing and improving, so the task of increasing the accuracy of the photogrammetry method using image segmentation is relevant.

Segmentation methods, such as region-based and watersheds, provide accurate segmentation only in the case of clear object boundaries and a significant difference between the color (brightness) of objects and the background [5, 6]. When obtaining images using digital video cameras, it is important that the contrast and brightness of the images are within the permissible range [7]. Otherwise (for example, when illuminated), some of the details in the image may be lost. It is possible to increase the accuracy of image segmentation using artificial intelligence [8, 9], in particular, using artificial neural networks (ANN) [10-12]. In image processing, convolutional neural networks (CNN) [13, 14] are among the most effective, the structure of which is adapted for processing two-dimensional signals.

Therefore, the paper proposes to perform image segmentation using convolutional neural networks YOLO (You Only Look Once) [15, 16]. Such neural networks allow for object detection in images, as well as their segmentation. According to the principle of operation of YOLO neural networks, objects in images are detected in one stage. An important advantage of neural networks is the ability to accurately detect and segment real images of objects that differ slightly from the background due to brightness, color and texture [17, 18]. Neural networks also allow for the selection of areas of objects of a certain class as segments. YOLO neural networks are pre-trained to detect 80 classes of objects, but can be further trained to detect objects of other classes.

The aim of the work is to increase the accuracy of constructing 3D models of objects using the photogrammetry method, which is implemented in the proposed information technology by pre-processing a series of images, namely, by segmenting images using neural networks YOLO and highlighting segments of certain classes as object masks. The object of research is the information technology for building 3D models using the photogrammetry method.

2. Related works

The basics of the photogrammetry method are described in [1]. The photogrammetry method is based on mathematical models that describe the geometry of the video camera, the spatial arrangement of the objects under study, etc. Both special and household cameras are used to obtain photographs of objects. Based on the analysis of a series of photographs, the three-dimensional coordinates of points on its surface are determined by the photogrammetry method.

In work [19] the creation of three-dimensional digital models of heritage sites using laser scanning (by LIDAR) and photogrammetry methods is described. It is shown that using laser scanning, a triangulation grid of a 3D model with high accuracy and detailed geometry is obtained. However, in the case of laser scanning, hidden areas often occur, especially for objects of complex shape. Such gaps can be significantly reduced by combining the results of laser scanning and photogrammetry. The initial images for the photogrammetry method are obtained by terrestrial and unmanned aerial vehicle (UAV) photography. The combination of photogrammetry with laser scanning has given good results in constructing three-dimensional models of historical buildings with complex architectural and decorative details. High-resolution photographs not only increase the accuracy of 3D models, but also improve the quality of textures. At the same time, it is important to obtain images of the object under study from all angles, which is possible when using a UAV.

Ways to improve the quality of models constructed using the 3D reconstruction method are considered in the study [20]. The importance of obtaining high-quality 3D models for building information modeling and city management is described. It is shown that obtaining a series of images of buildings using a UAV for the photogrammetry method is economically beneficial. Possible defects of models for buildings of complex shape, in which some parts may overlap. To improve the quality of building reconstruction, the study proposes a new method of planning the UAV trajectory, which involves obtaining images of all surface fragments from different viewpoints. Model construction errors were determined through the exact spatial coordinates x, y and z of checkpoints and the coordinates of these points obtained by the photogrammetry method. It is proposed to increase the accuracy of models by adding viewpoints for areas where larger construction errors are observed.

The modern neural network model SSG2 (Semantic Segmentation Generation 2) for semantic image segmentation is considered in work [21]. The peculiarity of the model is that it does not operate on separate images, but integrates the results of several images (observations). The SSG2 model uses a basic ANN with dual-encoder, single-decoder base network augmented with a sequence model. Due to the consideration of several similar images, the accuracy of their segmentation is increased. The resulting segmented images are used in the photogrammetry method and in the analysis of object images.

The initial images for the photogrammetry method can be obtained as individual frames of the video stream. Obtaining important frames (Keyshot) from the video stream is described in research [22]. The proposed algorithm allows obtaining important frames even from long video sequences.

Detection and segmentation of object images using neural networks YOLO (version 8) is described in research [15]. Neural networks YOLO are pre-trained on the COCO (Common Objects in Context) dataset, which contains over 100,000 images with labeled objects (80 classes). However, if necessary, YOLO networks can be trained on their own datasets. After training, the YOLO network selects the object under study in the image as a segment. The advantages of YOLO networks include high image processing speed and accurate selection of object boundaries, even of complex shapes. YOLO allows to detect objects not only in individual images, but also in real-time video. The accuracy of segmentation by YOLO networks in most cases exceeds the accuracy of known methods of region-based and watersheds [6].

The analysis of the reviewed publications confirms the relevance of research related to increasing the accuracy of the photogrammetry method using image segmentation by neural networks YOLO.

3. Proposed information technology for building three-dimensional models of objects

An information technology for constructing three-dimensional models of objects using the photogrammetry method is proposed, which provides an increase in the accuracy of constructing 3D models using image segmentation by neural networks YOLO. The information technology uses hardware (video cameras, computers and network equipment) and software (a Python program has been developed for image segmentation by neural networks YOLO and the formation of masked images, the 3DF Zephyr program for constructing 3D models using the photogrammetry method).

The initial images are written to the array *img*(i, k, c), where i = 0, ..., M-1; k = 0, ..., N-1, c = 0,..., Q_C; M, N are image height and width (in pixels); Q_C = 3 is number of color channels (red, green, blue) [6].

The process of constructing a 3D model includes the following steps:

1. Reading a series of Q_{Im} photos *img* of the object from the video camera.
2. Segmentation of images *img* by neural networks YOLO.
3. Extraction of masks for segments of the object in the image *img*.
4. Calculation of a scaled image of masks *mask_sc* based on the initial masks, where the scale of *mask_sc* corresponds to the scale of the initial image *img*.
5. Dilation of *mask_sc* with the number of iterations H_D, resulting in the calculation of a dilated image *mask_scD*; image dilation [6] is designed to increase the transition region between the object and the background.
6. Filtering *mask_scD* with a Gaussian filter kernel (standard deviation $\sigma_{HD} = H_D/3$, resulting in the calculation of the image *mask_scDG.*
7. Obtaining a masked image *img_mask* by multiplying the initial image *img* by the *mask_scDG* for all color channels.
8. Forming a set of initial *img* and masked *img_mask* images.
9. Building a three-dimensional model of the object by photogrammetry based on a set of images using the 3DF Zephyr program, saving the 3D model to files (.zep, .ply, Obj formats).

Steps #2-#7 are implemented using the developed software "YOLO_mask_24" in Python. Let's consider in more detail the individual steps of building the 3D model.

Step #2 consists of segmenting the initial images *img* with a pre-trained neural network YOLOv8 (one of the latest versions of YOLO) [15, 16]. The neural network for object segmentation, in addition to the object class and the probability (confidence) of its detection, returns the coordinates of the object center (x, y) in the image, its width and height, and *masks* for the object segments in the image. It is possible to detect several objects in the image at the same time, the network YOLO returns a segment for each detected object.

Step #6 consists of filtering the *mask_scD* by convolving it with a Gaussian filter kernel w_G (size $M_w \times N_w$ elements) with standard deviation $\sigma_{HD} = H_D/3$ (H_D is the width of the transition region between the object and the background, which is formed as a result of dilation) [6]. The filtered image *mask_scDG* with smoothed edges is calculated by the formula

$$mask\_scDG(i,k,c) = \sum_{m=0}^{Mw-1} \sum_{n=0}^{Nw-1} \sum_{c=0}^{Qc-1} mask\_scD(i_1,k_1,c) \cdot w_G(m,n)\,, \tag{1}$$

where i = 0, ..., M-1; k = 0, ..., N-1; c = 0,..., Q_C; $i_1 = i - m + m_C$; $k_1 = k - n - n_C$;

M, N are image height and width (in pixels);

m_C is center of the filter kernel in height; n_C is center of the filter kernel in width.

Smoothing the mask edges in the image *mask_scDG* provides a smooth transition between the object and the background.

The two-dimensional Gaussian function with the standard deviation σ_{HD} is used as the filter kernel w_G, which is described by the formula:

$$w_G(m,n) = \frac{1}{\sigma_{HD}\sqrt{2\pi}} exp\left(\frac{-((m-m_C)^2 + (n-n_C)^2)}{2\sigma_{HD}^2}\right) \quad (2)$$

where $m = 0, ..., M_w\text{-}1$; $n = 0, ..., N_w\text{-}1$;

m_C is the center of the filter kernel in height; n_C is the center of the filter kernel in width.

The sizes of the filter kernel are calculated taking into account the 3σ rule for the normal distribution: $M_w = 6 \cdot \sigma_{HD}$, $N_w = 6 \cdot \sigma_{HD}$.

Step #9 consists in constructing a three-dimensional model of the object using the photogrammetry method in the 3DF Zephyr program [2]. The construction of a 3D model of objects is performed on the basis of a set of Q_{IR} images, which are obtained from different angles using one or more video cameras. The positions of keypoints (for example, object boundaries or corners) are determined on the obtained images. Image matching consists in matching their keypoints. To construct the model, Q_{IS} images (from the set of Q_{IR} images) are used, for which satisfactory matching with other images by keypoints is obtained. Images that are poorly matched with others are not used for 3D reconstruction (as this may lead to model defects).

4. Software implementation of object mask selection

The software «YOLO_mask_24» for masking images of objects was developed in Python in the Google Colab cloud platform using the Jupyter Notebook and the computer vision library "OpenCV" [23, 24]. Image segmentation was performed by a convolutional neural network YOLOv8 [15] using the "ultralytics" library [17], from which the YOLO class is imported. Neural networks YOLO were used, pre-trained on the COCO training set [25].

A medium-sized neural network model for object detection is created with the command: model = YOLO("yolov8m.pt"), and a model for image segmentation is created with the command: model_seg = YOLO("yolov8m-seg.pt"). To increase the accuracy of segmentation (if necessary), the size of the neural network model can be increased, but this reduces the performance.

The method «predict» of the "yolov8m-seg.pt" model receives the initial image *img* as an input parameter, and returns the result in the form of an Ultralytics YOLO *Results* object, which contains information about the predicted masks, coordinates of the bounding boxes, classes and probabilities. The program leaves the bounding boxes only for objects of the given class.

The result of method «predict» is a list of ultralytics.engine.results.Results objects, where each object corresponds to one input image. The segmentation results are masks of the segments in the image and boxes objects, which contain:

- *cls* – object class number.
- *xyxy* – coordinates of the rectangular bounding boxes of the segments in the format [x_1, y_1, x_2, y_2].
- *conf* – object recognition probabilities (Confidence).

The calculation of the scaled image of the *mask_sc* based on the initial *masks* is performed using the bicubic interpolation method, namely the «resize» function of the OpenCV library. The scale of the *mask_sc* is equal to the scale of the initial image *img*, which allows spatially selecting arbitrary segments.

As a result, the dilation of the *mask_sc* calculates the dilated image of the mask mask_scD. The dilation is performed by the distance H_D (in pixels), which creates a transition area between the object and the background. The dilation is performed programmatically by the function «binary_dilation» of the «scipy» library.

The filtering of the *mask_scD* with a Gaussian filter kernel is performed by the function «gaussian_filter» of the «scipy» library, as a result of which the image *mask_scDG* with a smooth transition between the object mask and the background is calculated.

5. Results of constructing three-dimensional models of objects

5.1. Results of constructing a three-dimensional model of an object without masking

In accordance with the proposed information technology (section 3), three-dimensional models of an object (bench) were constructed. For this purpose, a set of Q_{Im} images *img* of the object was obtained using a video camera (Figure 1). The position and orientation of the video camera were chosen so as to obtain an image of the object from all angles.

Let us consider an example of constructing a three-dimensional model of a bench in the integrated 3DF Zephyr software package based on Q_{IR} = Q_{Im} images without masking. The construction of a 3D model (3D Reconstruction) consists in creating:

1. Sparse Point Cloud.
2. Dense Point Cloud.
3. Meshes of polygons.
4. Textured Meshes.

In the process of determining the alignment with neighboring images and the orientation of the cameras, satisfactory values were obtained for all images, so the 3D model was built on the basis of Q_{IS} = Q_{Im} images. The coefficient of the relative number of images taken into account when building the model is equal to $k_I = Q_{IS}/Q_{IR} = 1$. A high value of Q_{IS} and k_I ensures a relatively high quality of construction of the front part of the bench (Figure 2), however, defects appeared in the rear part of the bench (part of the rear supports of the bench are missing) (Figure 3, Figure 4).

Figure 1: A fragment of the set of initial images of the object (the set contains Q_{Im} = 23 images).

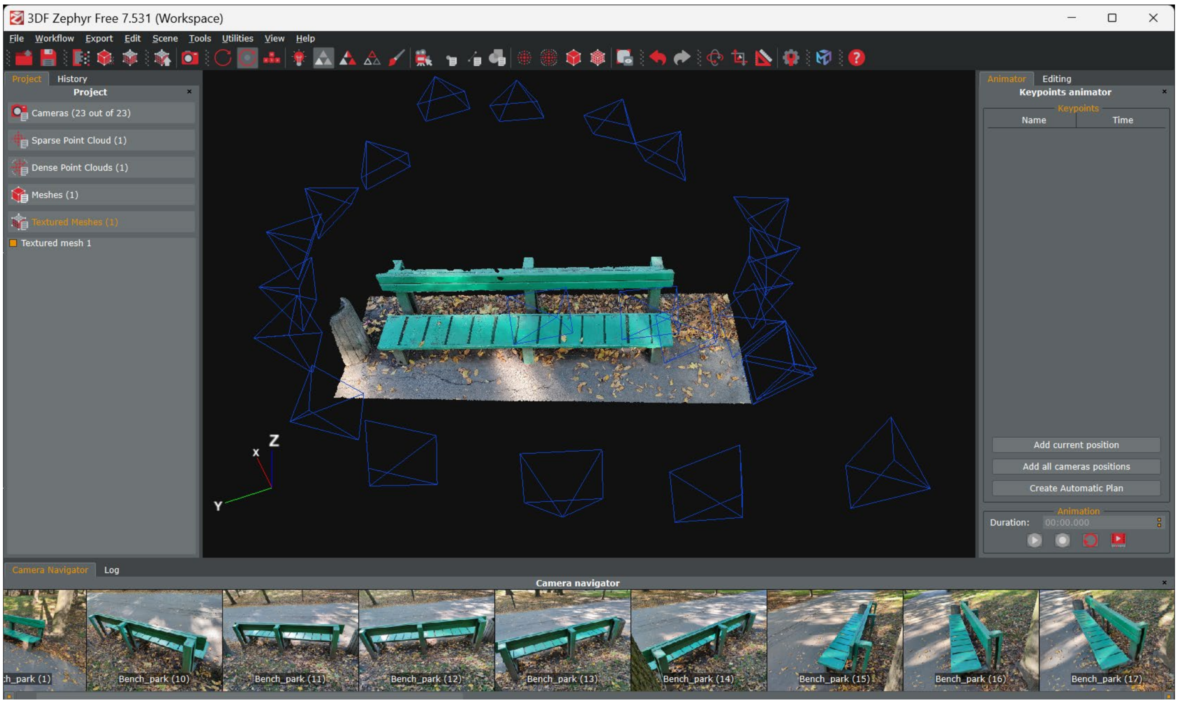


Figure 2: 3D model as a triangulation mesh (Points – 302686, Triangles – 395105) with textures after trimming excess details with a rectangular section, blue pyramids indicate the position of the camera at the time of image acquisition

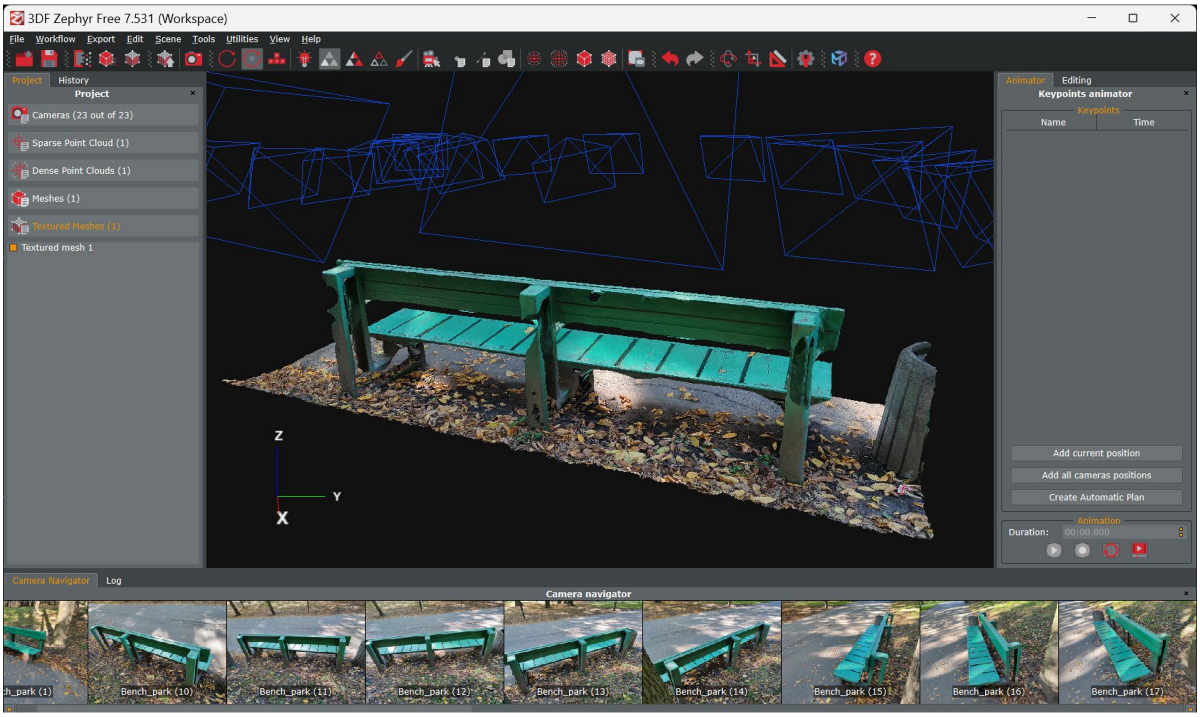


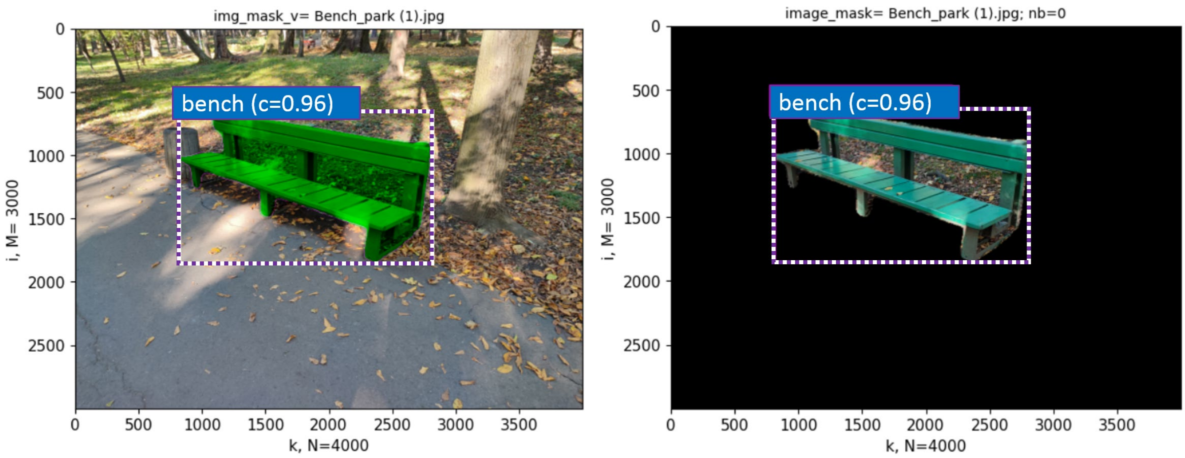
Figure 3: 3D model as a triangulation mesh with textures (back of the bench).

Figure 4: Fragments of the 3D model as a triangulation grid (right and rear parts of the bench), approximately 40% of the surface of the rear supports is damaged.

This result can be explained by the small number of images of the object (bench) and the negative impact of background objects that were captured by the frame. Therefore, to increase the accuracy of the model, the image set was masked.

5.2. Results of constructing a three-dimensional model of an object with masking

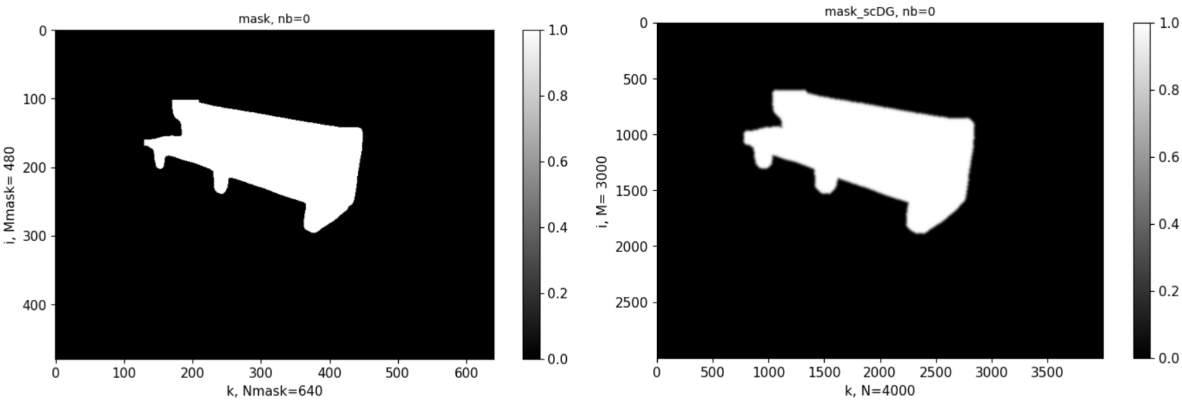
Let us consider an example of constructing a three-dimensional model of a bench in the integrated software package 3DF Zephyr based on $Q_{IR} = Q_{Im}$ images (Figure 1) after masking. Image masking was performed using the software «YOLO_mask_24». In each image, the neural network YOLO selects segments of objects, and for further processing, only segments of a given class (class «bench» with number 13) are used (Figure 5). Image masking consists in the fact that all areas of the image outside the frame of the studied object are painted over with a uniform color (black) (Figure 5b).



a) b)

Figure 5: Image of the object highlighted by a frame: a) initial image; b) masked image *img_mask*; *nb* is object number.

The image of the mask (Figure 6a) is scaled to the size of the original image (Figure 6b). Distortions (edge effects) appear on the clear borders of the mask, so around the mask it is possible to create a transition area between the object and the background using the dilation method (area width H_D = 30) and perform smoothing of the transition area with a Gaussian filter (Figure 6b). Multiplying the mask image by the original image results a masked image *img_mask* (Figure 7).



a) b)

Figure 6: Selection of the mask of the studied object: a) initial *mask*; b) *mask_scDG* after scaling, dilation to a width of H_D=30 pixels and smoothing with a Gaussian filter kernel with σ_{HD}=H_D/3 = = 10 pixels.

The images processed by masking (Figure 8) were used to build a three-dimensional model of the object using the photogrammetry method in the integrated software package 3DF Zephyr [5]. In the case of building a three-dimensional model only based on images after masking, some of the images are not taken into account by the program 3DF Zephyr (Q_{IS} = 17 images with Q_{IR} = 23, k_I = Q_{IS}/Q_{IR} = 17/23 = 0.739 were selected for building), since in this case it is more difficult to find common points between different images. As a result, this leads to deterioration in the quality of the model, so building a model only based on images after masking is impractical. This applies to both images with clear mask boundaries and images with blurred boundaries (Figure 7).

a) b)

Figure 7: Image fragments with the *img_mask*: a) with clear mask boundaries; b) with expanded and blurred mask boundaries.

Figure 8: A fragment of the image set of the *img_mask* object with a clear mask (the set contains Q_{Im} = 23 images).

A significantly better quality of the model is obtained when using simultaneously images without a mask (Figure 1) and images with a mask (Figure 8) for its construction. In this case, the model is built on the basis of $Q_{IR} = 2Q_{Im} = 46$ images.

At the same time, smaller distortions of the 3D model were obtained in the case of a mask with clear boundaries (the coefficient of the relative number of images selected for building the model, $k_I = Q_{IS}/Q_{IR} = 44/46 = 0.957$), since when using a mask with blurred boundaries, distortions occur in the transition areas and the accuracy of image matching in the photogrammetry method deteriorates. Therefore, the best result was obtained when using simultaneously images without a mask and images with a clear mask for building the 3D model (Figure 9, Figure 10, Figure 11).

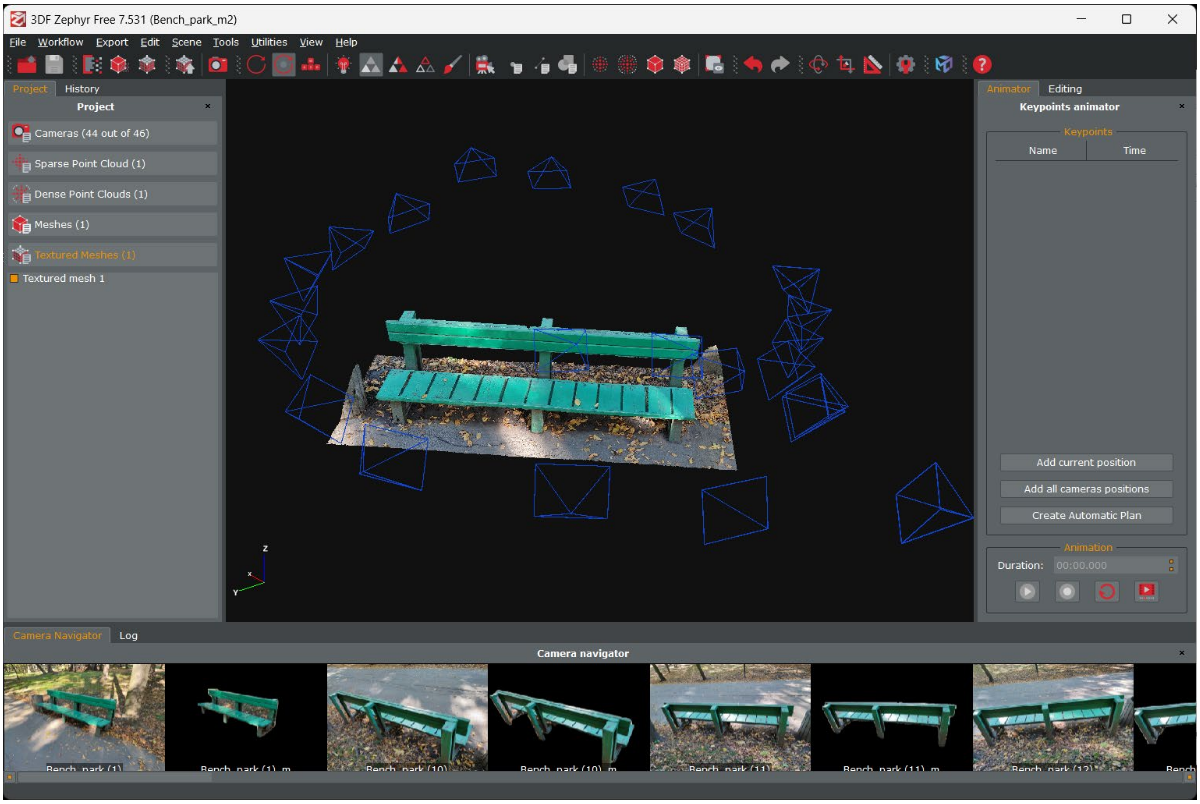


Figure 9: 3D model as a triangulated mesh (Points - 250 132, Triangles - 331 378) with textures after trimming excess details with a rectangular area; images without a mask and images with a clear mask were used to build the 3D model.

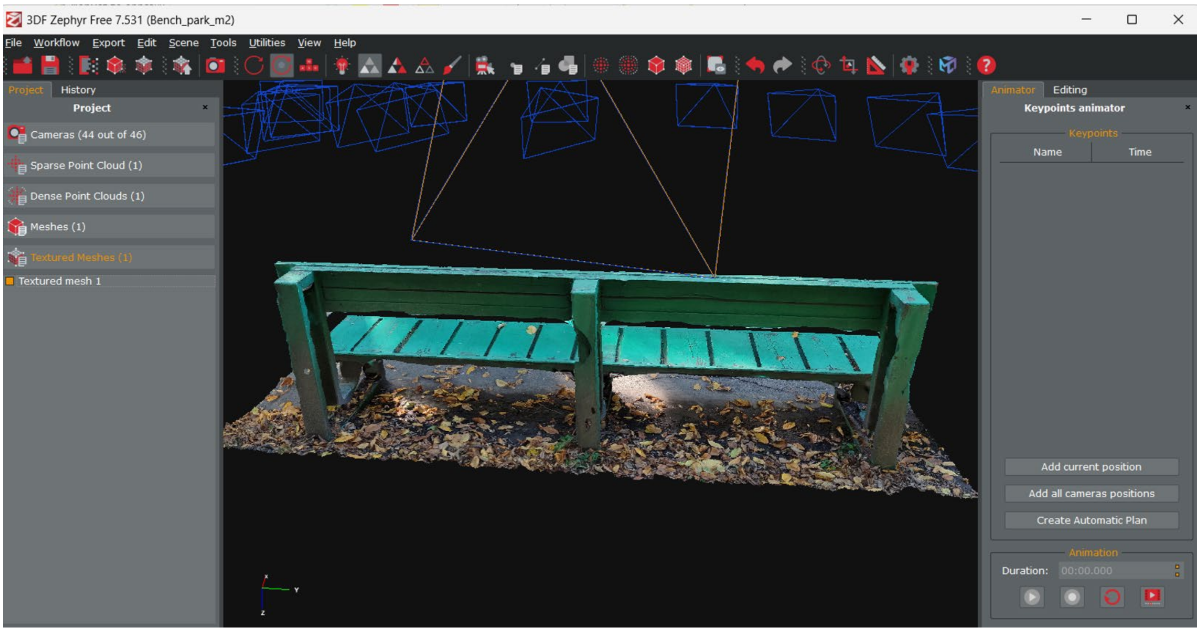


Figure 10: 3D model as a triangulation mesh with textures (back of the bench).

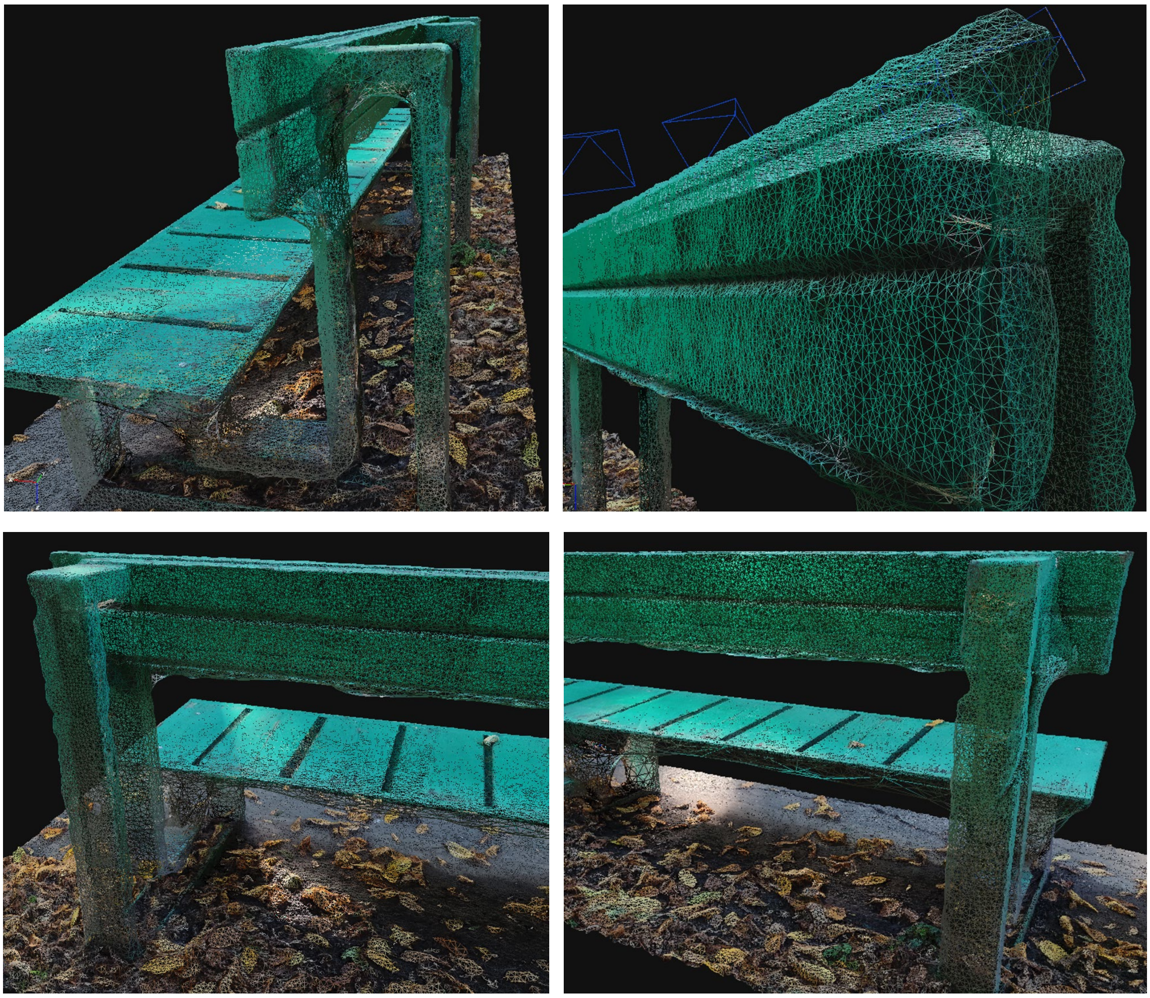


Figure 11: Fragments of the 3D model as a triangulation mesh (right and rear parts of the bench), the surfaces of the rear supports are practically not damaged.

As a result (Figure 11), high quality of all parts of the 3D model was obtained. Thus, by constructing an object model using simultaneously images without a mask and images with a mask, it was possible to reduce 3D model defects. Due to the absence of extraneous objects on the masked images, higher accuracy of constructing 3D models of objects is ensured. Images without a mask are required for better matching of different images with each other, since key points of not only the object, but also the background are used to match the images. The results of the research show that by simultaneously using images without a mask and images with a mask (obtained by segmenting the initial set of images), it is possible to increase the accuracy of constructing 3D models of objects.

6. Conclusion

An information technology for constructing three-dimensional models of objects using the photogrammetry method is proposed, which provides an increase in the accuracy of constructing 3D models due to automatic masking of images. The hardware of the technology consists of video cameras, computers and network equipment. The software of the technology consists of a program developed in the Python language for forming masked images, the 3DF Zephyr program for constructing three-dimensional models using the photogrammetry method. The software "YOLO_mask_24" was developed in Python on the cloud platform Google Colab, which is designed for segmentation, scaling, dilation, filtering, and masking of object images. Image segmentation is performed by a convolutional neural network YOLOv8.

Three-dimensional object models (benches) were constructed using the 3DF Zephyr program based on Q_{IR} = 23 images without masking using the photogrammetry method. All Q_{IS} = 23 images were used for construction. Sparse Point Cloud, Dense Point Cloud, Meshes and Textured Meshes were created. The created model contains 302686 Points, 395105 Triangles. The accuracy of most of the object model is satisfactory, however, defects appeared in the rear part of the model. Part of the rear supports of the bench is missing, approximately 40% of the surface of the rear supports is damaged.

It was found that building models using photogrammetry only on the basis of masked images leads to significant model defects, since in this case the images are poorly combined with each other at key points and only Q_{IS} = 17 images from Q_{IR} = 23 images (relative image number coefficient k_I = 0.739) are used to build the 3D model. This applies to both images with clear mask boundaries and images with blurred boundaries.

The novelty of the work is the simultaneous use of images of objects without masking and with masking to build their three-dimensional models using the photogrammetry method. To verify this approach, three-dimensional object models (benches) were built using the 3DF Zephyr program based on Q_{IR} = 46 images without masking and with masking. All Q_{IS} = 44 images were used for construction (relative image number coefficient k_I = 0.957). The created model contains 250132 Points, 331 378 Triangles. High accuracy was obtained for all parts of the 3D model.

The accuracy of constructing three-dimensional models of objects using the photogrammetry method has been improved, therefore the goal of the work has been achieved. It has been shown that the use of artificial neural networks YOLO allows to automate image masking for the photogrammetry method, increase the accuracy and speed of image processing, and obtain more accurate and detailed 3D models. The resulting 3D models of objects are used, in particular, in three-dimensional computer graphics, virtual and augmented reality systems, reverse engineering.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] E. Stylianidis, Measurements: Introduction to Photogrammetry. in: Photogrammetric Survey for the Recording and Documentation of Historic Buildings. Springer Tracts in Civil Engineering, Springer, Cham. 2020, pp. 139-195. doi: 10.1007/978-3-030-47310-5_6.

[2] 3DF Zephyr. The Complete Photogrammetry Solution, 2024, URL: https://www.3dflow.net.

[3] Mikeroyal. Photogrammetry-Guide, 2024, URL: https://github.com/mikeroyal/Photogrammetry-Guide?tab=readme-ov-file

[4] I. Nikolov. Python Libraries for Mesh, Point Cloud, and Data Visualization (Part 1), 2024, URL: https://towardsdatascience.com/python-libraries-for-mesh-and-point-cloud-visualization-part-1-daa2af36de30

[5] S. Palani, Principles of Digital Signal Processing, Springer Cham., 2022.

[6] R. Gonzalez, R. Woods, Digital image processing, Pearson/ Prentice Hall, New York, 2018.

[7] S. Balovsyak, Kh. Odaiska, O. Yakovenko, I.Iakovlieva, Adjusting the Brightness and Contrast parameters of digital video cameras using artificial neural networks, in: Proc. SPIE, Sixteenth International Conference on Correlation Optics 12938, 2024, pp. 129380I-1 - 129380I-4. doi: 10.1117/12.3009429.

[8] S. Russell, P. Norvig, Artificial Intelligence. A Modern Approach, Pearson Education, 2021.

[9] T. Hovorushchenko, V. Kysil, Selection of the artificial intelligence component for consultative and diagnostic information technology for glaucoma diagnosis, Computer Systems and Information Technologies 4 (2023) 87–90. doi: 10.31891/csit-2023-4-12.

[10] A. Geron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc., 2019.

[11] N. K. Manaswi, Deep Learning with Applications Using Python, Apress, India, 2018. doi: 10.1007/978-1-4842-3516-4.
[12] S. Balovsyak, I. Fodchuk, Kh. Odaiska, Yu. Roman, E. Zaitseva, Analysis of X-Ray Moiré Images Using Artificial Neural Networks, in: Proceedings of the 3nd International Workshop on Intelligent Information Technologies and Systems of Information Security (IntelITSIS 2022), March 23–25, 2022, ceur-ws.org, Khmelnytskyi, Ukraine, 2022, pp. 187-197.
[13] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects, IEEE Transactions on Neural Networks and Learning Systems 33 (12) (2022) 6999-7019. doi:10.1109/TNNLS.2021.3084827.
[14] S. Balovsyak, O. Kroitor, Kh. Odaiska, A.-B.M. Salem, S. Stets, Car Image Recognition using Convolutional Neural Network with EfficientNet Architecture, in: Proceedings of the 5th International Workshop on Intelligent Information Technologies & Systems of Information Security with CEUR-WS (IntelITSIS 2024), March 28, 2024, ceur-ws.org, Khmelnytskyi, Ukraine, 2024, pp. 182-195.
[15] How to Detect Objects in Images Using the YOLOv8 Neural Network, 2024, URL: https://www.freecodecamp.org/news/how-to-detect-objects-in-images-using-yolov8.
[16] Sik-Ho Tsang. Review: YOLOv8. Outperforms YOLOv5, YOLOv6 and YOLOv7, 2024, URL: https://sh-tsang.medium.com/review-yolov8-object-detection-5214fa105731
[17] Model Prediction with Ultralytics YOLO, 2024, URL: https://docs.ultralytics.com/modes/predict/#plotting-results
[18] Sik-Ho Tsang. Brief Explanation of YOLOv5, It Outperforms EfficientDet, 2024, URL: https://sh-tsang.medium.com/brief-review-yolov5-for-object-detection-84cc6c6a0e3a
[19] Z. Wu, P. Marais, H. Rüther, A UAV-based sparse viewpoint planning framework for detailed 3D modelling of cultural heritage monuments, ISPRS Journal of Photogrammetry and Remote Sensing 218 (B) (2024) 555-571. doi: 10.1016/j.isprsjprs.2024.10.028.
[20] Sh. Zhang, Ch. Liu, N. Haala, Guided by model quality: UAV path planning for complete and precise 3D reconstruction of complex buildings, International Journal of Applied Earth Observation and Geoinformation 127 (2024) 103667. doi: 10.1016/j.jag.2024.103667.
[21] F.I. Diakogiannis, S. Furby, P. Caccetta, X. Wu, R. Ibata, O. Hlinka, J. Taylor, SSG2: A new modeling paradigm for semantic segmentation, ISPRS Journal of Photogrammetry and Remote Sensing 215 (2024) 44-61. doi: 10.1016/j.isprsjprs.2024.06.011.
[22] A. Zarichkovyi, I.V. Stetsenko, Boundary Refinement via Zoom-In Algorithm for Keyshot Video Summarization of Long Sequences, in: Advances in Artificial Systems for Logistics Engineering III. ICAILE 2023. Lecture Notes on Data Engineering and Communications Technologies. Springer, Cham. 180, 2023, pp. 344–359. doi: 10.1007/978-3-031-36115-9_32.
[23] K. Kargin, Computer Vision Fundamentals and OpenCV Overview, 2024, URL: https://medium.com/mlearning-ai/computer-vision-fundamentals-and-opencv-overview-9a30fe94f0ce.
[24] G. Guillen, Digital Image Processing with Python and OpenCV, in: Sensor Projects with Raspberry Pi. Maker Innovations Series, Apress, Berkeley, CA, 2024, pp.105–147. doi: 10.1007/979-8-8688-0464-9_5.
[25] COCO (Common Objects in Context), 2024, URL: https://cocodataset.org/#home.