Improving the noise immunity of visual navigation algorithms based on the use of semantic descriptions of observed scenes

Dmitry S. Girenko [1] *

[1]MIRP-IS LLC, Dubna, Russian Federation - dima@webiceberg.com

Keywords: Technical Vision Systems, Visual navigation, Semantic description of scenes

Abstract:

The problem of estimating the coordinates of unmanned aerial vehicles (UAVs) using visual navigation in the absence of satellite navigation signals is considered. A camera is installed on the UAV, pointing towards the underlying surface, to assess the position by comparing the current images received on board with a reference image – a map of the area prepared in advance. The aim of the study is to increase the noise immunity and computational performance of visual navigation algorithms by switching from comparing bitmap images to comparing the content of observed scenes. In this case, the content of the scenes is presented in the form of semantic descriptions, including classes of objects, their attributes and the relationships between them. A technique for forming semantic descriptions of observed scenes based on the use of a neural network of the U-net architecture and computer vision algorithms is presented. Identification of scenes observed in the video camera with reference images is carried out using the Jaccard function. It is shown that the use of semantic descriptions increases the noise immunity and computational performance of UAV position estimation algorithms.

1. Introduction

Currently, small unmanned aerial vehicles (UAVs) are playing an increasingly important role in the world, navigation of which is mainly based on the use of a free-form inertial navigation system (INS) and Global Navigation Satellite System (GNSS) signals. The use of INS without correction by GNSS signals leads to an increase in errors in estimating the coordinates of the UAV over time.

Visual navigation methods are used to correct the operation of INS (Semenova, 2018) in the absence of GNSS signals (Geng and Chulin, 2017). One of the widely used correction methods for aircraft is orientation correction using the correlation extreme navigation system (CENS) (Beloglazov et al., 1985). The CENS is based on the idea of comparing the current image obtained from the UAV's video camera with a reference image (digital terrain map) stored in the on-board computer. The accuracy and noise immunity of such systems significantly depend on changes in the observation conditions of the current images. In addition, the need to compare high-dimensional bitmap images in real time places high demands on the performance of the onboard computer (Kim, 2001).

The purpose of this work is to increase the efficiency of visual navigation algorithms in terms of noise immunity and computational performance based on comparing semantic descriptions of current and reference images, instead of the traditionally used bitmaps.

2. An algorithm for visual navigation based on the semantic description of a scene

To achieve this goal, it is proposed to switch from algorithms based on the use of raster descriptions to algorithms that extract the content of the observed scene, i.e. to a semantic description. The semantic description of the observed scenes is understood as an enumeration of the classes of objects present in the image, a description of their features (shape, size, texture, etc.) and the relationships between them (relative location). It is important to note that the semantic description differs from the well-known semantic segmentation in that it is not the pixels of the image that are processed, but the semantic concepts behind them. Thus, the comparison of the current and reference images can only be based on comparing vectors with the listed objects, attributes of these objects, etc. In this case, the difference in the illumination of the images ceases to play a dominant role in the accuracy of the comparison algorithms.

Let's distinguish 3 stages of semantic description formation: objects in an image,features of objects, and relationships between objects. Figure 1 shows the UAV location detection algorithm based on a semantic description.

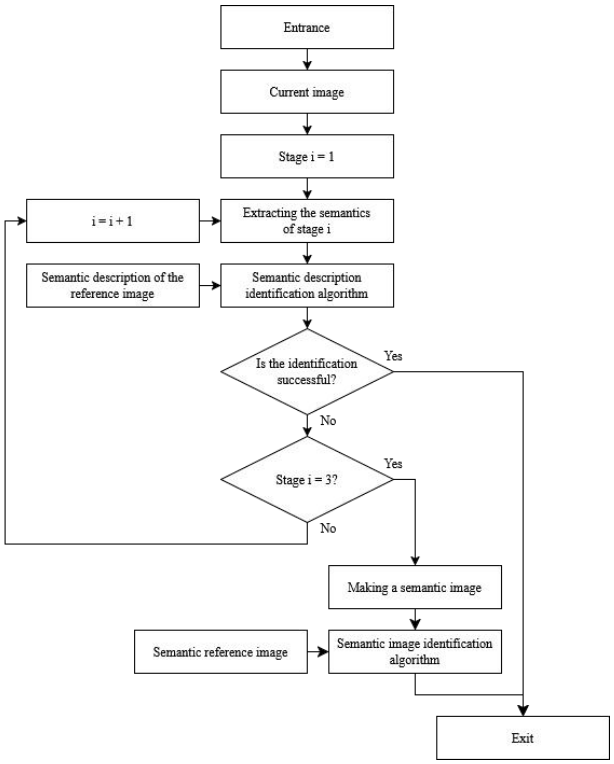

Figure. 1. An algorithm for determining the location of UAVs based on a semantic description

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W9-2025-111-2025 | © Author(s) 2025. CC BY 4.0 License.

Stage 1 - search for objects in the image. This stage is carried out by semantic segmentation by a neural network of the U-net architecture (Ronneberger et al., 2015). As a result of the neural network operation, each pixel of the segmented image contains the object class number. However, there are some errors in the segmented image: merging of different objects, splitting one object into several, the presence of unrecognized pixels inside the object, and uneven borders. These inaccuracies are eliminated using morphological image processing algorithms. The object classes found in the image are placed in a vector for further identification. If the description of the scene is not informative enough to unambiguously determine the location of the UAV, it is necessary to refine the description (go to the next stage), which, however, imposes additional costs.

Stage 2 - identification of the object's features. This stage primarily consists in forming a list of informative features. This article analyzes the shape of objects by comparing the number of points of the approximating contour of the object. The object class-shape pairs found in the image are placed in a vector for further identification.

Stage 3 - determining the relationships between objects. . This stage also requires the identification of those relationships that will be most informative during the flight mission. As an example, the ratio of an object's proximity to its neighbors is considered, whether the object is close, far away, or at an average distance. The pairs “object class - relation - object class” found in the image are placed in a vector for further identification.

If, after all the stages, no understanding has been reached about the location of the UAV, a semantic image identification algorithm is used - the result of semantic segmentation. Thus, the result of each stage is a vector containing information about the objects represented in the image. At the same time, it is possible to use the results of previous stages to refine the description of the scene in the following ones, such as in stages 2 and 3, the result of selecting objects in stage 1 is used. It is worth noting that at each stage it is possible to obtain not just a vector, but a full-fledged mask image containing pixel-by-pixel encoding of the found classes, attributes, and relationships. This can be useful at the stage of semantic image processing to increase sensitivity.

Since vectors and mask images obtained at the stages can be represented as binary images, it is proposed to use paired objective functions (Kim, 2001). In the work (Kim et al., 2025), it was shown that using the Jaccard function in conjunction with semantic descriptions reduces the operating time by several times. The position of the UAV is determined by searching for the maximum estimate of the paired Jaccard function for the current and reference images.

$$K = \frac{a}{a+b+c}. \quad (1)$$

where a - the number of matching elements of the vector or mask image;

b, c - the number of mismatched elements corresponding to the “missing target” and “false alarm" errors.

It is worth noting that the use of this function is allowed in two versions.

In the first variant, pixel-by-pixel estimation is performed when processing raster or semantic images, which has increased accuracy, but also greater computational complexity.

The second option allows for a semantic description of the scenes, represented as vectors. This option has less computational complexity and can be used for a preliminary assessment of the possible location of the UAV.

To reduce the amount of calculations, it is proposed to use the following approaches:

1.Using the INS error model, determine the area of probable UAV locations and perform a search in sequence from the most likely location to the least likely. At the same time, this area also serves as the boundary for the reference image that is being searched.
2. Preliminary analysis of the most informative objects, features, and relationships to cut off areas of the reference image without these features.
3. Using a paired function according to the second option with the cutting off of guaranteed unsuitable areas.

The general algorithm of the proposed solution is shown in Figure 2.

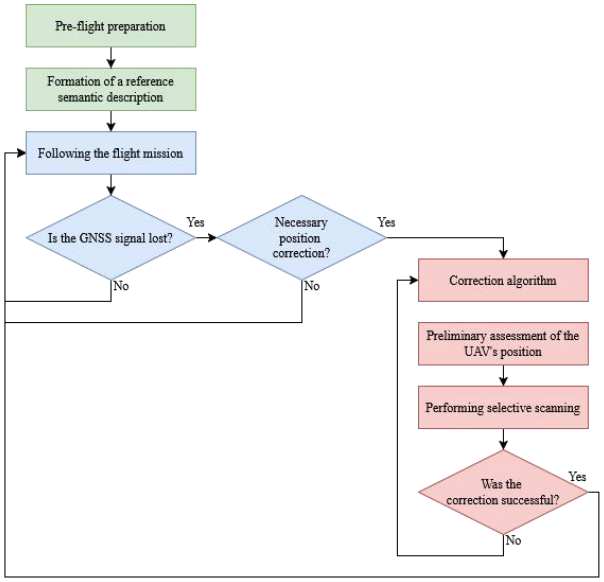

Figure. 2. The general algorithm of the proposed solution

Let's highlight the main points of the general algorithm:

1. During pre-flight preparation, a semantic description of the terrain of the proposed route is formed on the reference images. A table with a description of the scenes is being formed. The most informative features are highlighted.
2.During flight, when the correction condition is reached, the semantic description is displayed on the received current image.
3. A preliminary search is performed for the most likely locations of the UAV.
4. Selective scanning is performed in the found potential positions of the UAV by pixel-by-pixel calculation of the paired function.
5. If there is an unambiguous estimate of the paired function, its maximum value corresponds to the desired location of the UAV, the correction is completed. If several possible locations exceed the confidence threshold, additional information must

be entered, and no adjustments are made. However, this estimate will be used to speed up the next iteration.

3. Experiments and Results

As an example, an autonomous UAV flight (without using GNSS) with a video camera and a visual navigation system installed on board is considered (Figure 3). At time A, the UAV knows its position with sufficient accuracy to complete the target task. After a certain time, the UAV appears in zone B, however, the location of the UAV is known with an accuracy of ANN. Since the ANN error can be represented by a random variable with a normal distribution, the region B can be described by a normal probability density. Thus, it is necessary to find a more accurate location of the UAV within the boundaries of area B.

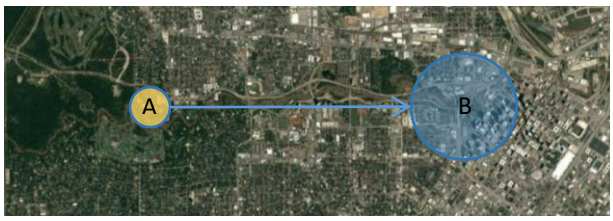

Figure. 3. Moving the drone from area A to area B

Let area B correspond to the area of the terrain map shown in Figure 4.

Figure. 4. Example of a section of an area map

In this case, we will assume that the observed image in the camera of the UAV has the form of Figure 5.

Figure. 5. The observed current image

Let's use the proposed algorithm for determining the location of a UAV based on a semantic description.

To implement the first stage, a neural network of the U-net architecture was trained based on a sample of semantic_segmentation_satellite_imagery (Alchimowicz, 2022). Figure 6 shows the markup for the area shown in Figure 4, and Figure 7 shows the result of a trained neural network. Roofs of houses are marked in red, grass is green, forest is dark green, water surfaces are blue, roads are white, and the rest of the classes are in shades of gray.

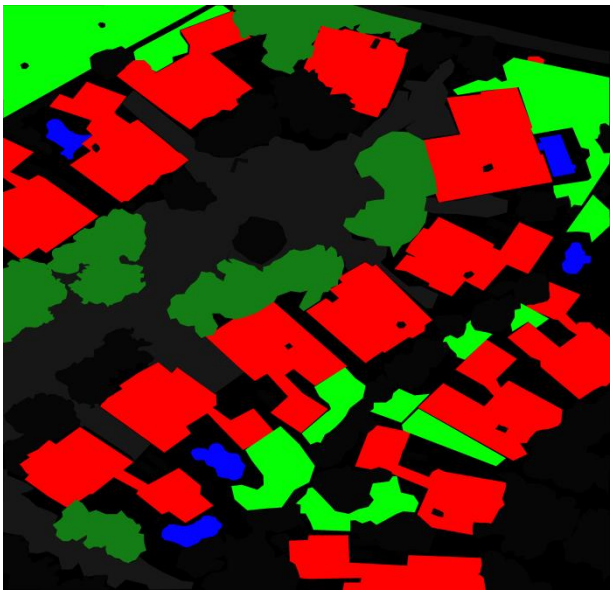

Figure. 6. Markup for the reference image

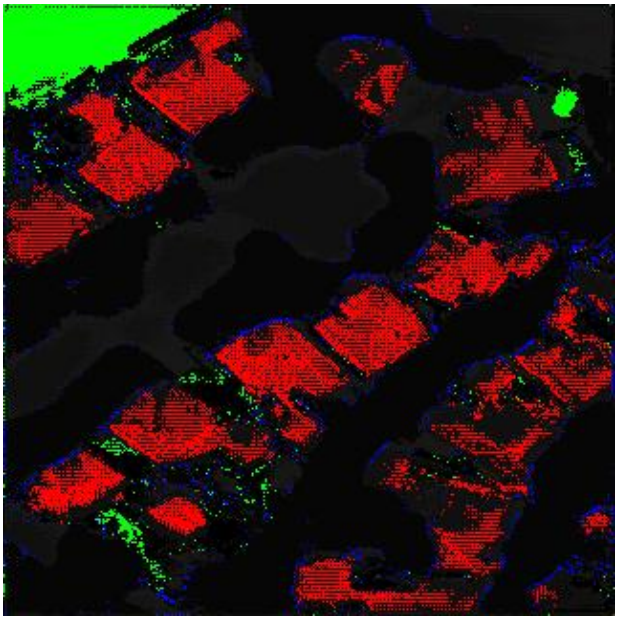

Figure. 7. The result of the neural network operation

As can be seen in Figure 7, the result of the neural network contains many inaccuracies, as each pixel is evaluated individually. The result of post-processing using morphological processing is shown in Figure 8.

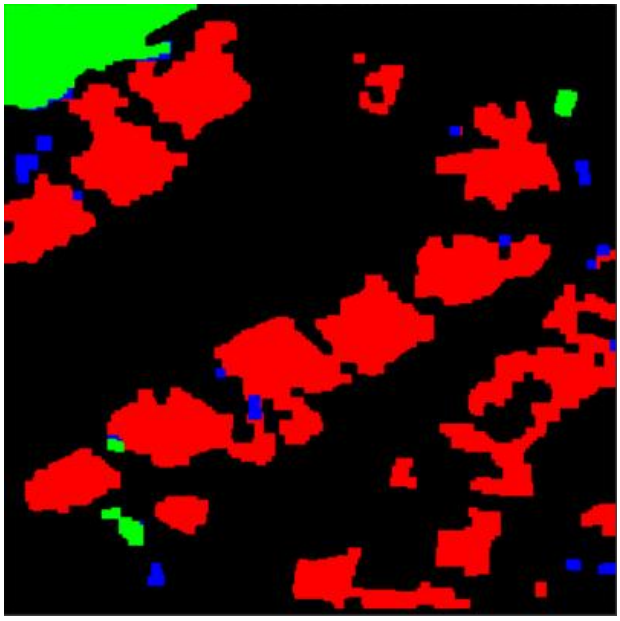

Figure. 8. The result of the neural network after morphological processing

Since the trained neural network best detects grass and road pixels, we will use only objects of these classes. Let's select individual objects in the current image using the contour search algorithm and get the following list:

<House, House, Grass, House>

This list of objects is a semantic description of the scene, which is used at the first stage of determining the location of the UAV. Since the UAV has an estimate of its height and the base of reference images was also obtained at a known height, we will scan the area, highlighting the semantic description and comparing it with the semantic description of the current image using the Jaccard function. It is worth noting that in this paper the order of the objects in the vector is not important. The result of the identification operation is shown in Figure 9.

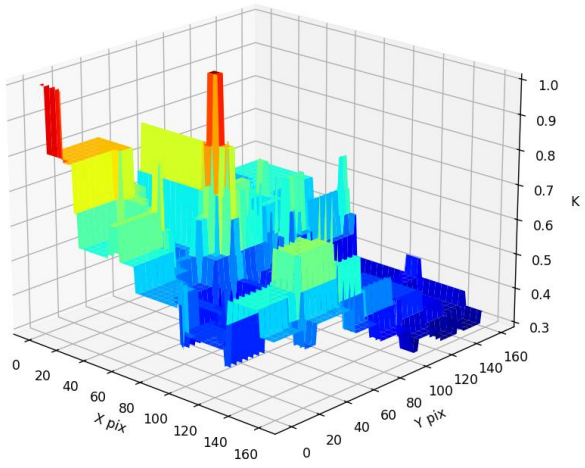

Figure. 9. The result of the identification of the first stage

As can be seen in the graph below, there are two most likely, equally probable locations of the UAV. A transition to the second stage is required for clarification.

As mentioned earlier, the second step is to evaluate the shape of the object. The shape is estimated by approximating the contour and then counting the number of vertices. The result is written as a sequence of class-number of vertices.

<House-4, House-8, Grass-5, House-9>

By analogy with the first stage, the Jaccard function is considered. Figure 10 shows the result of the second stage.

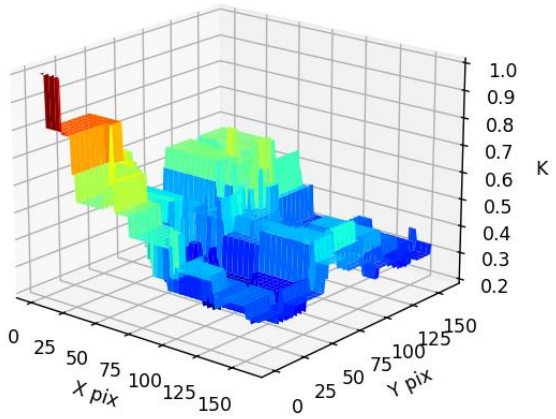

Figure. 10. The result of the identification of the second stage

Analyzing graph 10, we note that the second stage made it possible to unambiguously determine the position of the UAV. The task is completed, however, let's consider the following steps of the proposed algorithm.

The third stage consists of determining the distance between objects. The distance is estimated based on the found contours of objects, and then translated into a qualitative description using a threshold function: close, far, at an average distance. It is worth noting that sampling thresholds can be obtained during pre-flight training by analyzing the terrain map. The result is written as a sequence class-range-class.

<House-close-House, House-close-House, House-close-Grass, House-average-House, House-average-Grass, House-average-Grass>

By analogy with the previous steps, the Jaccard function is considered. Figure 11 shows the result of the third stage.

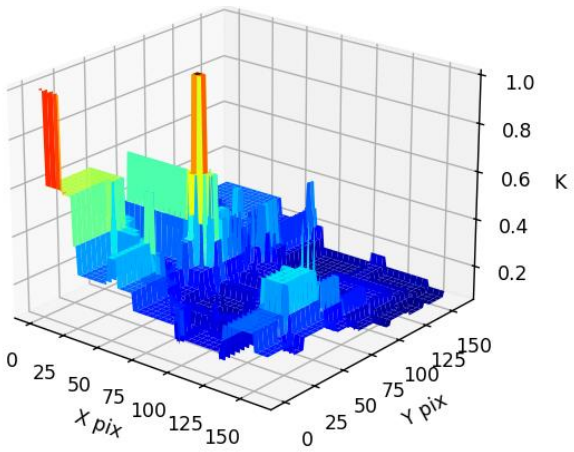

Figure. 11. The result of the identification of the third stage

Analyzing graph 11, we note that the result is extremely similar to the result of the first stage, however, due to the input of additional information in the form of relationships between objects, secondary peak values have reduced their probability.

Let's consider the result of calculating the Jaccard function for all pixels of the segmented image in Figure 12.

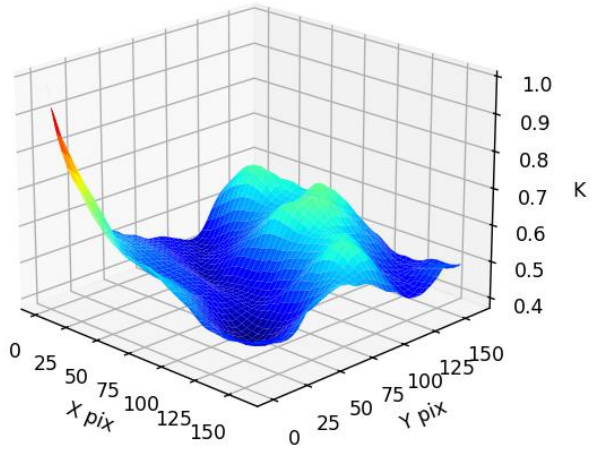

Figure. 12. The result of a pixel-by-pixel comparison by the Jaccard function

After analyzing graph 12, we note that in this case, the UAV was definitely able to determine its location. Now let's compare it with the result of the mutual normalized correlation function, the result of which is shown in Figure 13.

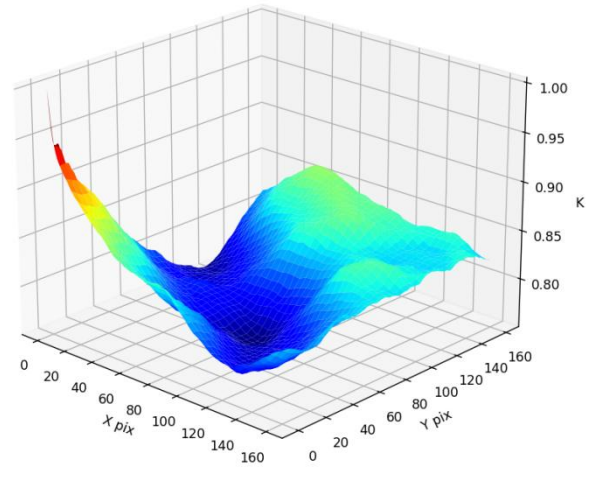

Figure. 13. The result of the correlation function

Comparing graphs 12 and 13, we note that the correlation function also coped with the task of detecting the location of the UAV, but it has a much higher probability of secondary peaks.

Let us proceed to the analysis of the noise immunity of the proposed approach in comparison with the mutual normalized correlation function. As an interference, consider the change in image brightness and the noise of the current image with Gaussian noise. The comparison results are shown in Figures 14 and 15.

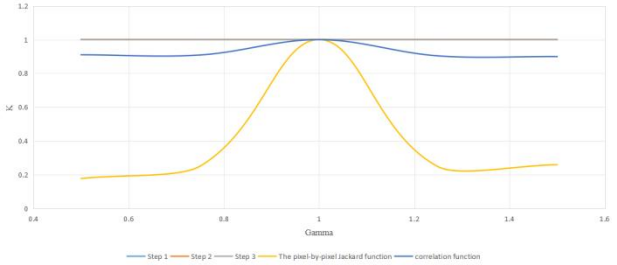

Figure. 14. The dependence of the similarity of the current and reference images on the brightness change of the current image

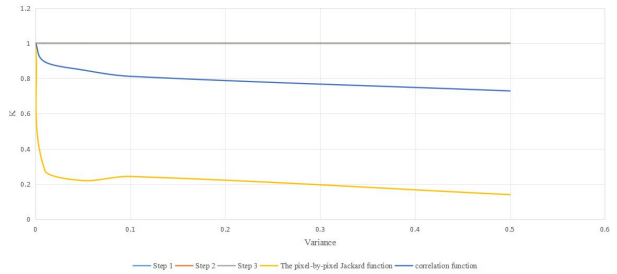

Figure. 15. The dependence of the similarity of the current and reference images on the Gaussian noise dispersion

After analyzing graphs 14 and 15, we can come to the following conclusions: stages 1, 2 and 3 of the proposed algorithm showed complete noise immunity. The results of these stages have not changed much. First of all, although the neural network has lost accuracy due to noise, it has had virtually no effect on the semantic description. However, we note that there may be cases when the loss of accuracy of the neural network will be critical for the algorithms of semantic description formation.

Although pixel-by-pixel processing of a segmented image using the Jaccard function dramatically loses its degree of similarity, this algorithm allows you to limit the range of possible UAV positions. Figure 16 shows an example of operation with Variance = 0.5

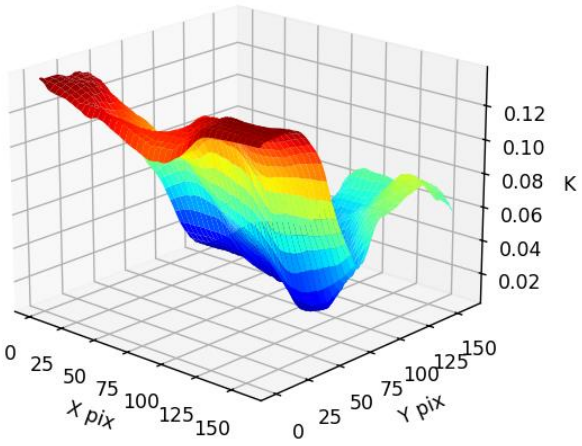

Figure. 16. The result of a pixel-by-pixel comparison by the Jaccard function when exposed to Gaussian noise with Variance = 0.5

The correlation function, despite its great degree of similarity, has the same problem as the pixel-by-pixel analysis of the Jaccard function - the generation of secondary peaks, similar in amplitude to the true location of the UAV (Figure 17).

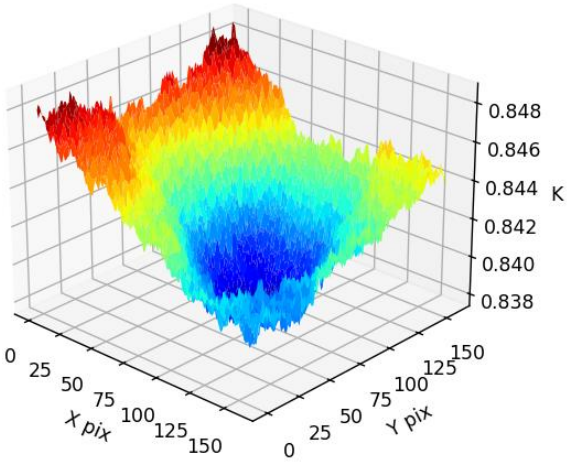

Figure. 17. The result of the correlation function when exposed to Gaussian noise with Variance = 0.5

Thus, it is shown that the proposed approach, namely the use of a semantic description, makes it possible to increase noise immunity. It is worth noting that the neural network underlying the production of a semantic image can be adapted to various influences by adding images to the dataset after applying such influences.

In conclusion, we will give the operating time figures for the various stages of the algorithm in comparison with the mutually normalized correlation function. The current image was 160 by 160 pixels, while the reference image was 320 by 320 pixels. The neural network took 0.5 seconds to run. The operation time of stage 1 took 0.701219 seconds. Stage 2 took 2.719419 seconds to complete. Stage 3 took 15.595586 seconds to complete. The operation time of the Jaccard algorithm applied to pixel-by-pixel analysis took 0.549907 seconds. The operation time of the cross-normalized correlation function took 59.722952 seconds.

Thus, in the worst case, the proposed algorithm takes 20.066131 seconds instead of 59.722952 seconds occupied by the correlation function.

4. Conclusions

1. In this paper, it is proposed to carry out visual navigation of UAVs based on semantic descriptions of scenes.
2. An algorithm for step-by-step refinement of the semantic description of scenes based on the use of a neural network of the U-net architecture and computer vision algorithms is presented.
3. Scenes are identified using the Jaccard function.
4. It is shown that the use of semantic descriptions increases the noise immunity of algorithms for estimating the position of UAVs.
5. The proposed solution allows to increase computing performance by more than 2 times.

5. Acknowledgements

The work was carried out within the framework of the state assignment of the Ministry of Science and Higher Education of Russia, topic No. FSFF-2024-0001.

6. References

1. Alchimowicz J. ,2022, semantic_segmentation_satellite_imagery. figshare. Collection. https://doi.org/10.6084/m9.figshare.c.6026765.v1.

2. Beloglazov I., Dzhadzhgava G., Chigin G., 1985, Fundamentals of navigation in geophysical fields. Moscow: Nauka.

3. Geng K., Chulin, A., 2017, UAV Navigation Algorithm Based on Improved Algorithm of Simultaneous Localization and Mapping with Adaptive Local Range of Observations. Herald of the Bauman Moscow State Technical University. Series Instrument Engineering, https://doi.org/10.18698/0236-3933-2017-3-76-94.

4. Kim N., 2001, Image processing and analysis in vision systems. Moscow : Publishing House of MAI, 84-94.

5. Kim N., Bodunkov N., Girenko D., Lyapin N., Udalova N., 2025, VISUAL NAVIGATION OF UNMANNED AERIAL VEHICLES USING SEMANTIC TERRAIN DESCRIPTIONS. Izvestiya SFedU. Engineering Sciences №2 (244), 256-268.

6. Ronneberger O., Fischer P., Brox T., 2015, U-Net: Convolutional Networks for Biomedical Image Segmentation, https://doi.org/10.48550/arXiv.1505.04597.

7. Semenova L., 2018, Modern methods of navigation of unmanned aerial vehicles, Science and education today No. 4 (27), 6-8.