Intelligent System for Visual Navigation

Victor Sineglazov
National Aviation University
Kyiv, Ukraine
svm@nau.edu.ua

Vitaliy Ischenko
National Aviation University
Kyiv, Ukraine
IschenkoVitaly@gmail.com

Abstract—**In this article it is considered the problem of algorithmic maintenance for visual navigation system. It is analyzed the characteristic features of different methods of intelligent system synthesis for visual navigation. The main stages and algorithms of work visual navigation system are encouraged to develop algorithmic maintenance for intellectual visual navigation system based on convolution neural networks.**

Keywords—unmanned aerial vehicle; visual navigation system; local features; computer vision; pixel coordinates; gps coordinates; global position; local position; roll; pitch; yaw; current image; previous image; landmarks

I. INTRODUCTION

One of the promising directions of improvement of navigation systems of aircraft is the use of machine vision technology.

At the same time the central problem – the development and implementation of landmarks image processing algorithms that provide reliable automatic binding coordinate ground scenes with limited power the onboard calculator.

II. FORMULATION OF THE PROBLEM

It is known a priori information about the presence of objects (targets) on a given area of the D, having previously unknown features and characteristics of the UAV mounted on the camcorder, which made scanning the earth's surface. There are the initial coordinates of the start point of the UAV (X_0, Y_0) and landmarks in the vicinity of this point (within 15 m) ..

Requirements:
- perform pre-treatment to remove video noise (to compensate for different lighting to images to filter noise;
- perform contrasting video.
- detect landmarks (descriptors) and highlight their features;
- determine their coordinates;
- based on the analysis of image transformation (comparing adjacent images to release them before certain features corresponding to the detected ground reference) define a pair of the same points in the use of the selected comparison method, compensating for the possible conversion of the image;
- the displacement of pairs of landmarks calculate the current position of the UAV.

III. REVIEW OF VISUAL NAVIGATION

Currently commercially available navigation system visual Visual Navigation System (Guidance) company DJI.

This system is composed of: a powerful processor, five pairs of ultrasonic transducers and 9 cameras , 8 of which are arranged around the perimeter in the horizontal plane and the one – is directed vertically downwards.

The system provides a navigation accuracy in the centimeter range. During the flight, the UAV speed, highly accurate stereo vision algorithms provide positioning information on any terrain. Due to of weak ultrasonic sensors working height functioning of the system up to 20 m.

Additionally, the system has the ability to measure airspeed, and integrated with an INS

IV. SYSTEM REQUIREMENTS TO VISUAL NAVIGATION

As the requirements for visual navigation system, should be following.

1) Finding of well-distinguishable points on the images:
- "features";
- "local feature points";
- "characteristic points".

2) Efficient mapping of points (feature matching):
- determining compliance with the starting points of the image and of the current;
- search conversion and matching found points.

3) Search and comparison of landmarks must be provided on the different types of textures of the earth's surface.

4) The original image may be subject to geometric transformations: rotate, scale, affine, affine geometry and change the brightness of the algorithm so visual navigation system must have the following invariant with respect to changes in the image:
- invariance to rotation;
- invariant to displacement;
- invariance to changes in scale;

- invariant to changes in brightness;

- invariant to changes in camera position point.

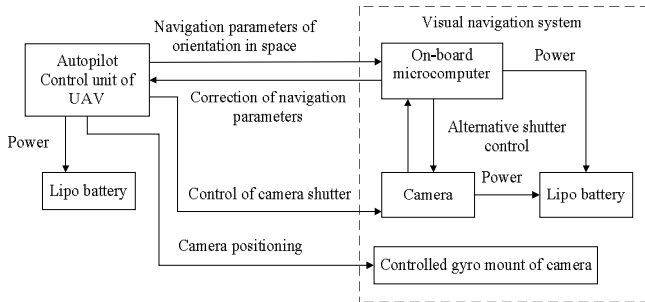In accordance with the requirements of a block diagram of a visual navigation system is shown in Fig. 1.

Fig. 1. Block diagram of the system of visual navigation.

Using a neural network implemented on-board micro-computer, will reduce the necessary computing power implementing parallel algorithms calculated visual navigation system and ensure greater efficiency.

V. ALGORITHMIC SUPPORT THE INTELLECTUAL SYSTEM VISUAL NAVIGATION

In general, the structure of algorithmic support of intellectual visual navigation system can be represented as follows.

1) The filtering algorithm of noise in the video image.
2) Algoritm segmentation (contrast) video.
3) Algorithm for finding local features in the image.
4) Algorithm for constructing and filtering pairs points (robust methods).
5) Algorithm for determining the coordinates of landmarks.
6) Algorithm for determining the coordinates of the current UAV.

Consider the need for, the description and characteristics of each of the algorithms listed above.

Appointment filtering algorithm is clear from its name. The following filters can be used to solve this problem:

- line filter;

- Wiener Filter;

- Gaussian filter;

- median filter;

- extreme filter;

- Kuwahara filter.

To solve the problem of segmentation is used a number of approaches, depending on the representation of the image, because:

- Burns method;

- Ekvelys method;

- Nibleks method;

- method and Janowitz Brukshteyna;

- methods based on clustering (for example, *k*-means, mean shift);

- using histogram techniques;

- methods sprawl areas (active contour);

- methods of graph cut;

- watershed method;

- algorithm for adaptive binarization.

Known algorithms for finding local features that allow find reference points in images: SIFT (Scale Invariant Feature), SURF (Speeded Up Robust Features), MSER.

Unfortunately, these algorithms have the disadvantages.

The algorithm SIFT:

- not all of the data points and their descriptions will meet the requirements that will affect future decision image matching problem;

- the algorithm does not work in the following cases;

- different lighting conditions (day / night);

- the object has a reflecting surface (usually vehicles, mirrors) expressed by 3-D structure, like structures (as a result, there are "true" non-compliance), the angle of the camera for a variety of very different images.

The algorithm SURF:

- in image processing does not stand out from the background object (features may be inside the object, and in the background, on the points of the object border and background);

- the algorithm does not work well: to objects of simple form and without a pronounced texture (in such objects, the method is not likely to find the singular points, points will be found either on the boundary of the object with the background, or even only on background)

MSER algorithm has the following disadvantages:

- not fully invariant to scale (scale normalization occurs depending on the size of the selected areas.

- Use sparseness function that requires a contrast in an image.

With this in mind, in this paper to solve the problem of finding landmarks are encouraged to use a convolution neural network.

A further object of visual navigation system is to identify and filter pairs points belonging landmarks found in the use of convolution neural network.

Under a pair of points is implied from the set point $N1$ $M1$ (previous image) and the corresponding point of the set $N2$ $M2$ (the current image). When comparing the points of the sets $M1$

and *M*2 wrong pair can be defined, which must be removed by applying filtering using robust or no robust methods.

Before determining the pairs of key points necessary to determine the model of the image geometric transformation (similarity (shift, scale, rotation), affine transformation, projective transformation of the plane (homography).

To solve the problem of constructing a pair of key points between the two images, there are several methods that are implemented in the library FLANN (Fast Approximate Nearest Neighbors).

The method constructs a fast search structure from a set of features using the specified algorithm with specified parameters. Parameter is a reference to one of the following classes. IndexParams descendants:

- ***linearIndexParams*** when passing an object of this type, the index will perform a linear, brute-force search;

- ***kDTreeIndexParams*** when passing an object of this type the index constructed will consist of a set of randomized kd-trees which will be searched in parallel;

- ***kMeansIndexParams*** when passing an object of this type the index constructed will be a hierarchical k-means tree;

- ***compositeIndexParams*** when using a parameters object of this type the index created combines the randomized kd-trees and the hierarchical *k*-means tree;

- ***lshIndexParams*** when using a parameters object of this type the index created uses multi-probe LSH (by Multi-Probe LSH:Efficient Indexing for High-Dimensional Similarity Search by Qin Lv, William Josephson, Zhe Wang, Moses Charikar, Kai Li., Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB). Vienna, Austria. September 2007);

- ***autotunedIndexParams*** when passing an object of this type the index created is automatically tuned to offer the best performance, by choosing the optimal index type (randomized kd-trees, hierarchical kmeans, linear) and parameters for the dataset provided.

After building a pairs (n, n') necessary to keep them in an array or file and filter steam that are outliers.

Methods for solving the problem of filtration steam points are divided into robust (emissions do not affect the resulting model) and no robust (volatile emissions). Therefore it is expedient to use a robust method that is M-SAC, which is a hybrid method, *M*-estimation and the method of RANSAC (Random Sample Consensus).

The filtering algorithm is given by the emission of pairs:

- find a model conversion points from the previous image of the current *N* and *N* '(as a model take camera shake), We estimate the model using the method of M-SAC (do a search pattern in the form of a shift);

- conduct emission filter for a better model.

To assess the current position of the UAV introduce the concept of pixel coordinates of significant points (determined by the method M-SAC) in the image, and the calculated GPS coordinates of these points (latitude, longitude).

The algorithm for determining the coordinates of the current UAV has the form:

- pixel scale score in meters for the current altitude;

- with respect to the original image at the point of takeoff UAV central coordinates of which are known, expect the GPS coordinates found and filtered reference points (assuming the data point relative to the central destination point image coordinates);

- for a particular pixel displacement of significant points in the current picture on the previous (the coordinates of which are known to us) to calculate the coordinates of the current points (apart from their destination);

- Of the known pixel coordinates of the center of the current picture (X_n, Y_{ts}), we calculate the GPS coordinates (the same coordinate of the current UAV), with respect to the distance from the current pixel relevant points calculated in the previous step.

To calculate the coordinates of significant points on the reference image at the point of take-off, it is necessary to determine the pixel distance from the image center, the coordinates of which are known to us, to the filtered by M-SAC points.

The pixel distance (offset) d between the points $A(x_1, y_1)$ and $B(x_2, y_2)$ in the plane defined by the equation:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

the resulting distance is equal to the number of pixels which should be multiplied by the current scale in meters camera matrix at a certain height.

The tangent of the angle between the segment and the positive direction of the axis Ox is defined by the formula (this angle is measured from the axis *Ox* counterclockwise).

$$\mathrm{tg}\varphi = \frac{y_2 - y_1}{x_2 - x_1}.$$

Defined by this formula is the tangent slope of the line.

After determining distance to the point and angle, it is necessary to recalculate the coordinates of latitude and longitude, considering their destination relative to the central point of the image using the following formula:

Destination point given distance and bearing from start point. Given a start point, initial bearing, and distance, this will calculate the destination point and final bearing travelling along a (shortest distance) great circle arc.

$$\varphi_2 = \mathrm{asin}(\sin \varphi_1 \cdot \cos\delta + \cos\varphi_1 \cdot \sin \delta \cdot \cos\theta),$$

$$\lambda_2 = \lambda_1 + \text{atan2}(\sin\theta \cdot \sin\delta \cdot \cos\varphi_1, \cos\delta - \sin\varphi_1 \cdot \sin\varphi_2)$$

where, φ is latitude, λ is longitude, θ is the bearing (clockwise from north), δ is the angular distance d/R; d being the distance travelled, R the earth's radius.

Algorithm for UAV autonomous navigation on the calculated current GPS coordinates, has the form:

- reading of the nearest target point of the flight plan with the autopilot memory;
- conversion rate and pitch to point to possible changes in altitude, stabilizing the angular orientation of the UAV.

Once we have found and counted the displacement of pairs of key points necessary to determine the current position of the UAV and the course to the next point of the mission flight plan.

To solve navigation tasks autonomous flight, the following assumptions:

Each picture has a certain set of properties:

- gps coordinates of the images (they may not be, in this case, the first picture is taken as the starting point, and all other attached to it);
- known UAV initial height (when a GPS signal in the start point);
- known angle relative to the vertical line X;
- known angle relative to the horizontal line Y;
- known angle relative to the line Z.

When receiving the information from items 3, 4, 5 is not necessary to solve the system with respect to unknown 5:

$$f = \begin{vmatrix} 1 & \text{tg}\,\varphi & \dfrac{\text{tg}\,v}{\cos\tau} \\ X_l & Y_l & Z_l \\ X_r & Y_r & Z_r \end{vmatrix} (Z_l X_r - X_l Z_r)\,\text{tg}\,\varphi$$

$$+ \left(X_l Y_r - Y_l X_r\right)\dfrac{\text{tg}\,v}{\cos\tau} + Y_l Z_r - Z_l Y_r = 0,$$

where X_n, Y_n, Z_n, parameters are dependent on items 3, 4, 5, when determining the GPS coordinates of the image are determined by the relative orientation angles v, τ. And in the absence of coordinates solve systems of linear equations with two unknowns.

As a point of reference to the origin of the destination of the first image at the point of takeoff. Thereafter found landmarks in the image will be associated with this point and its GPS coordinates. Extrapolating the obtained coordinates of the images, and reading from the memory autopilot cars recorded the coordinates of points of the mission flight plan, get the coordinates of the target point and the current. Using haversine formula consider the distance to the target point through a great circle arc approximating the Earth sphere in meters.

$$a = \sin^2(\Delta\varphi/2) + \cos\varphi_1 \cos\varphi_2 \sin^2(\Delta\lambda/2),$$

$$c = 2\arctan 2\left(\sqrt{a}, \sqrt{(1-a)}\right),$$

$$d = Rc,$$

where φ_1 current longitude, φ_2 longitude of the target point of the mission is loaded in memory UAV autopilot, λ latitude, R radius of the Earth (the average radius 6.371 km).

Next, you need to calculate the rate and pitch at a target point for navigation Autopilot. In general, your current heading will vary as you follow a great circle path (orthodrome); the final heading will differ from the initial heading by varying degrees according to distance and latitude (if you were to go from say 35°N, 45°E ($\approx$ Baghdad) to 35°N, 135°E ($\approx$ Osaka), you would start on a heading of 60° and end up on a heading of 120°!). Initial bearing (sometimes referred to as forward azimuth) which if followed in a straight line along a great-circle arc will take you from the start point to the end point:

$$\theta = \text{atan2}(\sin\Delta\lambda \cdot \cos\varphi_2, \cos\varphi_1 \cdot \sin\varphi_2 - \sin\varphi_1 \cdot \cos\varphi_2 \cdot \cos\Delta\lambda),$$

where φ_1, λ_1 is the latitude and longitude of current point, φ_2, λ_2 the end point ($\Delta\lambda$ is the difference in longitude). Since atan2 returns values in the range $-\pi$ ... $+\pi$ (that is, $-180°$ ... $+180°$), to normalise the result to a compass bearing (in the range 0° ... 360°, with values transformed into the range 180° ... 360°), convert to degrees and then use $(\theta + 360)\% 360$, where % is (floating point) module.

For final bearing, simply take the initial bearing from the end point to the start point and reverse it (using $\theta = (\theta+180)\% 360$).

After calculating the navigation parameters of the current picture based on the already calculated data in the image $N - 1$, you need to send data to the autopilot feedback to refine the positioning. It is necessary to clarify the positioning of the UAV with a frequency of 5 Hz (as a GPS signal), if resources onboard calculator does not allow it to do then you must go to 2 Hz frequency (2 frames per second), but extrapolation of these parameters.

VI. CONCLUSION

It is analyzed different approaches to the design of visual navigation systems, such as detection of landmarks, selection of neighborhoods descriptor reference points and it is chosen the effective methods for the solution of these problems. Implementation of the system based on the convolution neural network provides parallel processing of rows and columns of pixels in the image which greatly reduces the time, the additional use of texture analysis and ensuring invariance discussed in the report provides a constant number of reference points on the ground of different information content and the fixed run-time search and matching landmarks.

REFERENCES

[1] D.G. Lowe, "Distinctive image features from scale- invariant keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91-110, 2004.

[2] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in Proc. IEEE Conf. Comput. Vis. Pattern Recog, 2007, pp. 1-8.

[3] J. Sivic and A. Zisserman, "Video Google: A text retrieval

approach to object matching in videos," in Proc. IEEE 9th Int. Conf. Comput. Vis., 2003, pp. 1470-1477.

[4] J. Hays and A. A. Efros, "Scene completion using millions of photographs," ACM Trans. Graph., vol. 26, p. 4, 2007.

[5] G. Shakhnarovich, P. Viola, and T. Darrell, "Fast pose estimation with parameter-sensitive hashing," in Proc. IEEE 9th Int. Conf. Comput. Vis., 2003, pp. 750-757

[6] A.C. Berg, T.L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," in Proc. IEEE CS Conf. Comput. Vis. Pattern Recog., 2005, vol. 1, pp. 26-33.

[7] A. Torralba, R. Fergus, and W.T. Freeman, "80 million tiny images: A large data set for nonparametric object and scene recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 11, pp. 1958-1970, Nov. 2008.

[8] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei Fei, "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2009, pp. 248-255.

[9] J. L. Bentley, "Multidimensional binary search trees used for associative searching," Commun. ACM, vol. 18, no. 9, pp. 509-517, 1975.

[10] Paul D. Groves, Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems. Second Edition (Artech House Remote Sensing Library) 2nd Edition, 2008, p. 503.

[11] D. Arthur and S. Vassilvitskii. k-means++: the advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pages 1027-1035. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 2007.