

A method of iterative image normalization for tasks of visual navigation of UAVs

M O Elantcev¹, I O Arkhipov¹ and R M Gafarov¹

¹Kalashnikov Izhevsk State Technical University, Russian Federation, Izhevsk, Studencheskaya 7, 426069

e-mail: emo-name@hotmail.com, aio1024@mail.ru

Abstract. The work deals with a method of eliminating the perspective distortion of an image acquired from an unmanned aerial vehicle (UAV) camera in order to transform it to match the parameters of the satellite image. The normalization is performed in one of the two ways. The first variant consists in the calculation of an image transformation matrix based on the camera position and orientation. The second variant is based on matching the current frame with the previous one. The matching results in the shift, rotation, and scale parameters that are used to obtain an initial set of pairs of corresponding keypoints. From this set four pairs are selected to calculate the perspective transformation matrix. This matrix is in turn used to obtain a new set of pairs of corresponding keypoints. The process is repeated while the number of the pairs in the new set exceeds the number in the current one. The accumulated transformation matrix is then multiplied by the transformation matrix obtained during the normalization of the previous frame. The final part presents the results of the method that show that the proposed method can improve the accuracy of the visual navigation system at low computational costs.

1. Introduction

This paper is concerned with the development of a method for image normalization that is used in a visual navigation system of unmanned aerial vehicles (UAVs). The navigation system being developed [1] matches images acquired from an UAV onboard camera with the satellite image of the flying area. The matching yields the shift, rotation, and scale values that can be used to transform the input image to the corresponding part of satellite image. Knowing the coordinates of the satellite image and the shift, rotation, and scale parameters of the current input image, it is possible to determine the coordinates of the UAV.

During the flight the tilt of the UAV is unstable, it changes because of weather conditions or when the UAV is making a pivot. It results in perspective distortions of images acquired from the onboard camera, even if a stabilized platform used. The distortions interfere with the correct matching of the images from the camera with the satellite image, which is usually represented in an orthogonal projection. To increase the likelihood of correct matching a method of image normalization for eliminating the perspective distortions is needed. The method should be computationally simple because it is supposed to run on low-power UAV processors.

There are several approaches to solve this problem. The first approach is to calculate the matrix of perspective transformation based on the external parameters of the camera obtained from the UAV

sensors, namely pitch, yaw, roll, and altitude. In [2], the distortion is compensated by determining the orientation of a virtual camera with its optical axis collinear to the normal vector of the shooting plane. This virtual camera provides us with an image without perspective distortions. Therefore, to eliminate the distortions of the input image we need to calculate the homography matrix between the images from the real and the virtual cameras. There are two difficulties with using this approach: the moments of shooting and of external parameters recording should be strictly synchronized, and the surface of the flying area should be flat.

The second approach is to find and analyze the location of the known structural elements in the input image. This approach assumes that each frame from the camera is processed independently. The paper [3] focuses on the distortion elimination for the task of bar code recognition. Its idea is to find four straight lines bounding the quadrilateral barcode area. These lines provide four corner points that are used to calculate the transformation matrix. This approach is difficult to apply to the task of normalizing the UAV images because the shapes of the objects in the images are usually unknown.

The third approach is to find the pairs of corresponding elements in the two images and use their positions to determine the transformation matrix of one image into another. Usually the methods search for the pairs of image points with characteristic local features known as keypoints. Depending on the task, different methods of keypoints extraction and matching can be used [4,5].

Under ideal conditions, four pairs of the keypoints are enough to determine the transformation matrix. However, in practice even the positions of the correctly matched points can be different because of noise. To obtain a more accurate solution a system of equations based on all the matched pairs is constructed. This system is usually solved by the least squares method so as to minimize the given error function (the DLT method [6]). In [7], it is proposed to perform this method iteratively. This method helps to increase the likelihood of image matching by means of eliminating the noise points.

The set of the pairs of corresponding points can also contain mismatched pairs. In this case the DLT method can yield a wrong result and one of the robust methods should be used, such as RANSAC [8, 9]. In this method four pairs of corresponding points are randomly selected among all the matched pairs. Using them the homography matrix is calculated, and then the algorithm checks how many pairs fit this transform (inliers). The process is repeated, and other random pairs are selected from the set of inliers until a solution is found that satisfies the specified number of pairs or the maximum number of iterations is exceeded. Another similar robust method is LMS. It estimates four current selected pairs of points not by the number of inliers but by the median distance among all the pairs of points [6].

The described robust methods require a large number of iterations; therefore, they are not effective on low-power processors. The RHO method [10] aims to solve this problem. Instead of randomly electing four pairs of points at each iteration, the method preliminarily evaluates each pair, selects the most promising ones and then corrects the estimation. This allows to obtain the result in significantly fewer iterations.

The fourth approach is to analyze the input image using a neural network, as it is demonstrated in [11]. It is noted that the advantage of this approach is the ability to work when there is a very small number of the keypoints or when the input images are blurred. However, this method is computationally complex and requires special hardware to work in real time, which makes it difficult to apply to the task of UAV visual navigation.

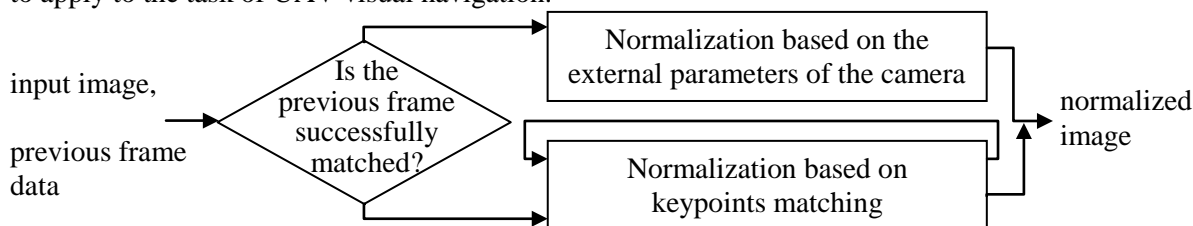


Figure 1. Simplified scheme of the proposed method.

The proposed method of iterative image normalization for tasks of visual navigation of UAVs combines the first and the third approaches. A simplified scheme of the method is shown in Figure 1. Normalization based on the external parameters of the camera (pitch, yaw, roll, and flight altitude) is used either when processing the initial camera frame or in the case where the previous frame has not been successfully matched with the satellite image or the current frame cannot be matched with the previous one. In other cases the iterative method of homography matrix calculation based on matching the keypoints of the previous and the current frames is used.

2. Description of the peculiarities of the task considered

The UAV flies along the given trajectory that contains both straight sections and turns. The considered flight altitude is 250-500m. There is a camera onboard the UAV that is directed strictly downwards. There is no stabilized platform on the UAV. The shooting speed of the camera is enough to provide at least 80% of overlap between the current and the previous frames. The parameters of focal length and the dimensions of the camera's sensor are known.

The UAV is equipped with an accelerometer and a pressure sensor that provide information about the UAV orientation (roll, pitch, and yaw) and the flight altitude. However, according to experiments on real images, the moments of shooting and of parameters recording are not exactly the same. Therefore, strong wind gusts cause inaccuracies in the recorded external parameters of the camera. In addition, the pressure sensor estimates the flight altitude only above sea level. When calculating the flight altitude above ground level the height map of the flying area should be taken into account.

The result of the normalization is the image transformed into the orthogonal projection so that later it can be matched with the satellite image.

3. Normalization based on the external parameters of the camera

The first algorithm of the proposed normalization method is based on the information on the camera orientation and position in space, namely the roll ρ , pitch τ , yaw γ , and flight altitude h_f . In addition, the input information includes the terrain height h_a at the shooting point and the reference altitude h_r at which the shooting scale equals the satellite image scale. The output information is the coefficients of the transformation matrix that eliminates the distortion. The algorithm consists of two stages: the calculation of the transformation coefficients disregarding the scale and the calculation of the scaling factor.

3.1. Calculation of the transformation coefficients disregarding the scale

The camera's field of view can be represented as a pyramid that intersects the earth's surface plane. To simplify the calculation, instead of rotating the pyramid, according to the camera orientation the earth's surface plane is rotated and placed at one of the base vertices of the pyramid, as Figure 2a shows.

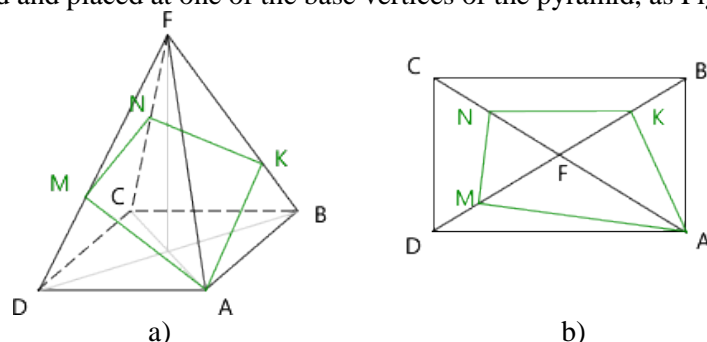


Figure 2. The pyramid of vision: (a) a view at an angle, MNK is the section with the earth's surface plane, (b) the top view.

The purpose of the first stage is to calculate the coordinates of the points M, N, K on the top view projection (Figure 2b). These coordinates can be used to calculate the resultant homography matrix T_a by solving an equation system constructed with the coordinates transformation equations of the pairs

of points: A-A, B-K, C-N, D-M. The transformation of each pair of points can be described by two equations [12]:

$$x' = \frac{t_{11}x + t_{12}y + t_{13}}{t_{31}x + t_{32}y + 1} \quad y' = \frac{t_{21}x + t_{22}y + t_{23}}{t_{31}x + t_{32}y + 1}, \quad (1)$$

where x and y are the coordinates of the point before the transformation, x' and y' are the coordinates of the point after the transformation, t_{11} , t_{12} , t_{13} , t_{21} , t_{22} , t_{23} , t_{31} и t_{32} are the coefficients of the transformation matrix T_a .

The shape of the pyramid depends on the size of the photo sensor and the focal length. The viewing angles are calculated as follows:

$$\alpha_x = 2 \operatorname{arctg} \frac{m_w}{2f_L} \quad \alpha_y = 2 \operatorname{arctg} \frac{m_h}{2f_L}, \quad (2)$$

where m_w and m_h are the dimensions of the photo sensor, f_L is the focal length. The vertex angles of the pyramid are calculated as follows:

$$\beta_x = \frac{m_w}{m_h} \sin \frac{\alpha_y}{2} \quad \beta_y = \frac{m_h}{m_w} \sin \frac{\alpha_x}{2} \quad (3)$$

The length of the edge of the pyramid l_e and the diagonal of the base of the pyramid l_d are calculated as follows:

$$l_e = \frac{m_h}{2 \sin(\beta_y/2)} \quad l_d = \sqrt{m_w^2 + m_h^2}. \quad (4)$$

Depending on the orientation of the camera, the earth's surface plane is placed at one of the four base vertices, as shown in Figure 3.

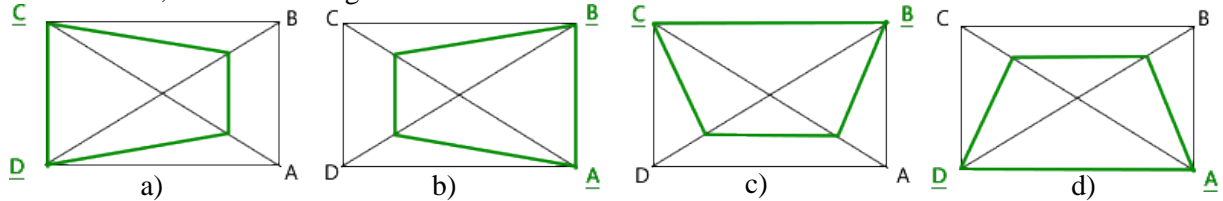


Figure 3. The location of the pyramid section with the earth's surface plane depending on the orientation of the camera: (a) $\tau = 0, \rho < 0$, (b) $\tau = 0, \rho > 0$, (c) $\tau > 0, \rho = 0$, (d) $\tau < 0, \rho = 0$.

The vertices are renamed according to the following rule: a vertex that coincides with a section corner is called A', then in anticlockwise direction: B', C', D'.

The coordinates of the vertex K can be calculated with the following formula:

$$d_k = m_w \frac{\sin \rho}{\cos(\beta_x/2 - \rho)}$$

$$K_x = \left| D'_x - d_k \frac{m_w}{2l_e} \right| \quad K_y = \left| D'_y - d_k \frac{m_h}{2l_e} \right| \quad (5)$$

The coordinates of the vertex M can be calculated with the following formula:

$$d_m = m_h \frac{\sin \tau}{\cos(\beta_y/2 - \tau)}$$

$$M_x = \left| B'_x - d_k \frac{m_w}{2l_e} \right| \quad M_y = \left| B'_y - d_k \frac{m_h}{2l_e} \right| \quad (6)$$

The coordinates of the vertex N can be calculated with the following formula:

$$d_{nk} = m_h - 2d_k \cos \frac{\pi - \beta_y}{2} - \sin \frac{\pi - \beta_y}{2} / \sin(\pi - \tau - \frac{\pi - \beta_y}{2})$$

$$d_{fk} = d_e - d_k \quad d_{nfb} = -2d_{fk} \cos \beta_y \quad (7)$$

$$d_{nf} = (\sqrt{d_{nfb}^2 - 4(d_{fk}^2 - d_{nk}^2)} - d_{nfb})/2$$

$$N_x = \left| \frac{C'_x - A'_x}{2} + d_{nf} \frac{m_w}{2l_e} \right| \quad N_y = \left| \frac{C'_y - A'_y}{2} + d_{nf} \frac{m_h}{2l_e} \right|$$

To align the image with the northward direction, matrix multiplication can be used:

$$T_n = T_a \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (8)$$

3.2. Calculation of the scale factor

Since the earth's surface plane was placed at one of the base vertices of the pyramid, the obtained transformation matrix disregards the original scale of the image. To calculate the scaling factor the following formula can be used:

$$d_{nd} = \sqrt{(N_x - A'_x)^2 + (N_y - A'_y)^2}$$

$$S_n = \frac{l_e d_{nd}}{d_{nf} l_d} \quad (9)$$

To transform the image acquired from the UAV camera to the scale of the satellite image, the transformation matrix should be adjusted using the following scale factor:

$$S_r = S_n \frac{h_r - h_a}{h_f - h_a} \quad (10)$$

3.3. Obtaining the result of the normalization

To obtain the final transformation matrix the following formula should be used:

$$T_r = T_n \begin{pmatrix} S_r & 0 & 0 \\ 0 & S_r & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (11)$$

Figure 4 shows the result of the method of image normalization based on the external parameters of the camera.

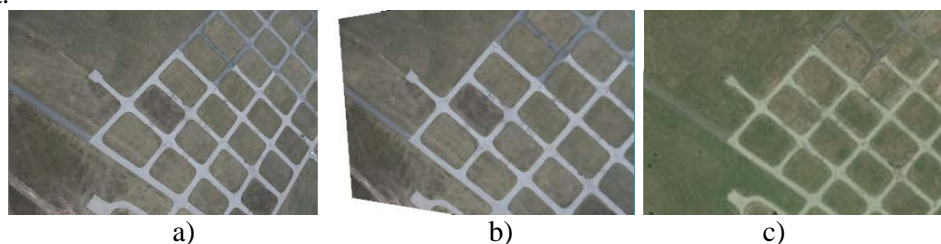


Figure 4. The result of image normalization based on the external parameters of the camera: (a) the original image, (b) the normalized image, (c) the corresponding part of the satellite image.

4. Normalization based on keypoints matching

The first algorithm of the proposed method eliminates the distortions of the image coarsely, because, as it is mentioned above, the moment of recording the external parameters of the camera not necessarily coincides with the moment of shooting. Moreover, it is rather difficult to accurately take into account the elevation difference of the underlying surface. Therefore, it is used only when the previous frame from the camera cannot be matched with the satellite image. In other cases the second algorithm is used that is based on matching the current frame with the previous frame already normalized.

4.1. Matching the current frame with the previous one

To reduce its complexity, the image matching algorithm analyzes not all the pixels but only the keypoints extracted with the method described in [13].

After the keypoint sets of the both images are found, they are used to determine the shift, rotation, and scale values between the two images using the statistical differentiation method [14]. These parameters form the following transformation matrix [15]:

$$H_m = \begin{pmatrix} s \cos \phi & -s \sin \phi & t_x \\ s \sin \phi & s \cos \phi & t_y \\ 0 & 0 & 1 \end{pmatrix}, \quad (12)$$

where t_x and t_y are the shift values along the X axis and the Y axis respectively, ϕ is the rotation value, s is the scale value.

4.2. Selecting pairs of corresponding points

After the transformation matrix is found, it is used to determine the pairs of corresponding points. Some point A of the current frame corresponds to some point B of the previous frame when after

applying the transformation matrix to the coordinates of the point A the distance between A and B does not exceed the specified threshold (1-2 pixels).

From the whole set of the pairs of corresponding points four pairs are selected that will be used to calculate the homography matrix (1). The choice is limited by the condition that no three points are collinear. The best case occurs when a quadrilateral of maximum area with as large angles as possible is found. However, the search for such a set of points is a computationally complex operation. The following approaches to selecting the points were tried:

- 1) the four points nearest to the corners of the image;
- 2) the four points most distant from the center of mass of the keypoints set and located in different quadrants relative to the center of mass;
- 3) the first point is the most distant from the center of mass, the second point is the most distant from point 1, the third is the point with the greatest sum of its distances from points 1 and 2, the fourth is the point with the greatest sum of its distances from points 1, 2 and 3;
- 4) a random selection of points.

If this algorithm finds less than four corresponding pairs, the normalization based on the external parameters of the camera is applied.

4.3. Calculation of the transformation matrix

After the four pairs of the corresponding keypoints are selected, the refined transformation matrix is calculated. This homography matrix matches the projection of the current frame with the projection of the previous frame already normalized.

Then the search for the pairs of corresponding points is repeated using the current frame warped according to the refined transformation matrix. From the found pairs a new set of four pairs is selected and a new refined matrix is calculated. The process is repeated as long as the number of the pairs of corresponding points increases. Experiments show that in most cases two iterations are enough to obtain the solution. The final matrix of the transformation of the original image into the normalized one is determined by the formula:

$$T_r = T_p \prod_{i=1}^n H_i \quad (13)$$

where T_p is the perspective transformation matrix for normalizing the previous frame, H_i are the transformation matrices obtained at each iteration, n is the number of iterations.

Since the previous frame is used to normalize the current frame, the normalization error can accumulate. However, this method is used as a part of the visual navigation system that implies that the next step is to match the normalized image with the satellite image. This matching eliminates the accumulated error.

5. Experiments

To test the proposed method, it has been integrated into the preprocessing stage of the UAV visual navigation system that is being developed. Then a series of experiments using real images of a UAV flight were performed. The trajectory of the UAV flight is shown in Figure 5 and consists of sections of two types: straight sections (270 frames) and turns (70 frames).



Figure 5. The trajectory used in the experiments. The positions of the UAV at the straight sections are marked with green dots. The positions of the UAV at the turns are marked with yellow dots. The black lines show the boundaries between the sections. The white lines connect the coordinates of the UAV to the corresponding centers of photographing.

The first experiment aims to choose the best algorithm for image normalization based on keypoints matching. The proposed iterative algorithm was tested (with the four variants of selecting the four pairs of points described above), as well as the methods implemented in the OpenCV 3.4.5 library: DLT, RANSCAC, LMS and RHO. To determine the speed of the methods, each input frame was processed 100 times and the average execution time was calculated. Processing was performed on an AMD FX-8320E 3.2 GHz processor without using a graphics accelerator.

Table 1. Comparison of the results of various methods of the homography matrix calculation.

Indicator	Section	Proposed method				DLT	RAN SAC	LMS	RHO
		variant 1 of pair selection	variant 2 of pair selection	variant 3 of pair selection	variant 4 of pair selection				
Average number of pairs of corresponding points before processing	straight sections	156.8							
	turns	123.2							
	whole trajectory	150.0							
Average number of pairs of corresponding points after processing	straight sections	173.5	171.5	172.1	158.9	175.0	177.5	178.3	178.2
	turns	136.7	133.6	135.0	125.6	138.7	141.3	142.6	142.0
	whole trajectory	165.8	163.7	164.5	152	167.5	170	170.9	170.7
Average number of iterations of the proposed algorithm	straight sections	1.9	1.8	1.8	1.2	-			
	turns	1.8	1.7	1.7	1.2	-			
	whole trajectory	1.9	1.8	1.8	1.2	-			
Average processing time per one frame (ms)	straight sections	0.06	0.05	0.07	0.03	0.33	0.52	2.72	0.313
	turns	0.05	0.04	0.06	0.03	0.31	0.57	2.70	0.24
	whole trajectory	0.06	0.03	0.06	0.03	0.33	0.53	2.72	0.30

Table 1 presents the processing results. The methods were evaluated by the average number of the pairs of corresponding points (greater values are better) and the average processing per one frame (lower values are better). According to the first criterion, the methods implemented in the OpenCV library were the best: RANSAC, LMS, RHO. They increased the number of the pairs of corresponding points by ~ 20 pairs (~ 13% of the initial quantity). However, the speed of processing is critical for the implementation of a UAV navigation system. According to this criterion, the proposed iterative method with the variants 1-3 for the pair selection was the best. It takes ~0.03–0.06ms for processing one frame, which is ~5–10 times faster than the best of the OpenCV methods, with that it increases the number of the pairs of corresponding points by ~15 pairs (~10%). The high speed of the proposed method is due to the fact that in most cases two iterations are enough to obtain the resulting homography matrix. Thus, for the second experiment according to the both criteria the following methods were chosen: the proposed iterative method with the first variant of the pair selection and the RHO method.

The second experiment aims to determine the effect of the integration of the normalization method into the UAV visual navigation system. There were three trials in the experiment: without normalization, with normalization using the proposed method, and using the RHO method. The first criterion was the number of the images from the UAV camera successfully matched with the satellite image. The second criterion was the average distance between the automatically calculated positions of the center of photographing and the manually marked ones. The results of the experiment are shown in table 2.

After analyzing the results of the experiment, we can conclude that the preliminary normalization of the image significantly improves the output of the UAV visual navigation system being developed. It increases the number of the successfully matched frames as well as the accuracy of the calculated coordinates of the UAV. The methods work well both in the straight sections of the flight trajectory and in the sections with turns. The result of the proposed method is not worse than the one of the RHO

method from the OpenCV library, and at that the proposed method performs much faster, which is very important for integrating it into a UAV visual navigation system.

Table 2. The effect of normalization methods on the output of the visual navigation system.

Indicator	Section	Without normalization	Proposed method with variant 1 of pair selection	RHO
Number of frames	straight sections	270		
	turns	70		
	whole trajectory	340		
Number of successfully matched frames	straight sections	226 (83.7%)	246 (91.1%)	246 (91.1%)
	turns	59 (84.3%)	64 (91.4%)	62 (88%)
	whole trajectory	285 (83.8%)	310 (91.1%)	308 (90.6%)
Average distance between the found center of photographing and the manually marked one (pixels)	straight sections	16.0	6.6	6.8
	turns	7.5	6.4	6.7
	whole trajectory	16.0	6.6	6.7

6. Conclusion

The proposed method of iterative image normalization successfully solves the task of image preprocessing for developing a UAV visual navigation system. The integration of this method into the navigation system improves the accuracy of matching the images acquired from the onboard camera with the satellite image and reduces the probability of the UAV loss.

Due to using the two normalization approaches the proposed method can eliminate perspective distortions resulting from both the camera orientation and the inclination of the underlying surface. At the same time, this method is computationally simple and can be integrated into the UAV navigation system, which should run on low-power processors.

7. References

- [1] Gafarov R M, Arkhipov I O, Korobeynikov A V and Elantsev M O 2017 Application of the method of statistical differentiation for the gridding of aerophoto image to a satellite image *Intellektualnye sistemy v proizvodstve* **15** 109-112 DOI: 10.22213/2410-9304-2017-2-109-112
- [2] Kholopov I S 2017 Projective distortion correction algorithm at low altitude photographing *Computer Optics* **41(2)** 284-290 DOI: 10.18287/0134-2452-2017-41-2-284-290
- [3] Boriskevich A A 2016 A robust method for normalization of barcode images scanned with a mobile device *Izvestija Nacionalnoj akademii nauk Belarusi. Serija fiziko-tehnicheskikh nauk* **4** 106-112
- [4] Adel E, Elmogy M and Elbakry H 2014 Image stitching based on feature extraction techniques: a survey *International Journal of Computer Applications* **88** 1-8
- [5] Goncharenko M O 2015 Comparative analysis of key points detection methods as a part of video segmentation problem *Bionics of Intelligense* **85** 90-94
- [6] Hartley R and Zisserman A 2003 *Multiple View Geometry in Computer Vision* (Cambridge: Cambridge university press)
- [7] Efimov A I and Novikov A I 2016 An algorithm for multistage projective transformation adjustment for image superimposition *Computer Optics* **40(2)** 258-265 DOI: 10.18287/2412-6179-2016-40-2-258-265
- [8] Agarwal A, Jawahar C V and Narayanan P J 2005 A survey of planar homography estimation techniques *Centre for Visual Information Technology International Institute of Information Technology*
- [9] Goshin Y V, Kotov A P and Fursov V A 2014 Two-stage formation of a spatial transformation for image matching *Computer Optics* **38(4)** 886-891 DOI: 10.18287/0134-2452-2014-38-4-886-891
- [10] Bazargani H, Bilaniuk O and Laganieri R 2015 A fast and robust homography scheme for real-

- time planar target detection *Journal of Real-Time Image Processing* **15** 739-758 DOI: 10.1007/s11554-015-0508-4
- [11] DeTone D, Malisiewicz T and Rabinovich A 2016 Deep image homography estimation *arXiv* (Preprint arXiv:1606.03798)
- [12] Glasbey C A and Mardia K V 1998 A review of image-warping methods *Journal of applied statistics* **25** 155-171
- [13] Arkhipov I O, Mishenkov M A, Shutov A A and Elantsev M O 2018 Algorithm of unmanned aerial vehicle's positioning using sequence of video stream frames *Intellektualnye sistemy v proizvodstve* **16** 66-69 DOI: 10.22213/2410-9304-2018-3-66-69
- [14] Gafarov M R, Gafarov R M and Poltanova E S 2015 Determination of geometrical images parameters using statistical differentiation *Intellektualnye sistemy v proizvodstve* **25** 41-44
- [15] Chekhlov D O and Ablameyko S V 2004 Normalization of images relating to perspective transformation based on geometric options *Informatika* **3** 67-76

Acknowledgments

The reported study was funded by Kalashnikov ISTU according to the research project № 09.06.01/18 GMM.