AERIAL IMAGES STITCHING ALGORITHM BASE ON GEOMETRIC ALIGNMENT OF FRAMES

Naing Linn Aung[1], Koldaev Victor D.[2], Kyaw Zaw Ye[3]

Department of Informatics and Computer Software System, National Research University of Electronic Technology, Bld.1, Shokin Square, Zelenograd, Moscow, 124498, Russia

Abstract. In this article, the synthesis algorithm of a continuous image of the earth's surface from individual overlapping frames is proposed. The algorithm based on determining the law of geometric alignment of images using the coordinates of feature points identified in the common areas of adjacent frames. Examples of practical use of an algorithm for the geometrical alignment of frames received by using a digital camera are given. The paper reflects the features by using the Harris algorithm for reliable identification of feature points on the analyzed images. The problems of determining the adequate mathematical models, allowing to carry out a qualitative combination of images are investigated.

Keywords: aerial images, image stitching, geometric alignment of a frame, feature points.

1. Introduction

For imaging, must find out a correspondence between the scene points and image points, and then need to determine what is depending on the brightness in each image point.

The brightness of the surface area of the object is proportional to the brightness of the scene. The proportionality factor depends on the properties of the optical system. Since only a finite number of measurements can be transmitted to a computer, required spatial discretization. Measurements are performed at the nodes of a square raster or grid. Then, the image is represented as a rectangular array of integers. For ease of combining images into a panorama require a certain number of images. The required number of images and the degree of overlap are determined by the field of view of the lens used. The most widespread frame overlay at 30 percent and the number of shots from 2 to 30 (circular panorama).

Formation of the panoramic image composed of the following steps [1]:

1) determination of the singular points of two neighboring images;

2) finding correspondences between the singular points;

3) calculating a transform matrix;

4) overlay images and obtaining outcome.

The aerial image is formed by a set of individual images overlapping each other $B_k = [b_k(x, y)]$, where $b_k(x, y)$ – pixel brightness with coordinates (x, y), $x = \overline{1, X}$; $y = \overline{1, Y}$; k – the frame number ($k = \overline{1, K}$). The technology, which is based on geometrical transforming frames in a cartographic system of coordinates according to the elements of internal and external orientation parameters of the recording device is used to form a continuous image of the shooting route from a set of images B_k. However, in some cases, such as an aerial image with a digital camera or malfunctions of navigation equipment, exterior orientation are absent. Then, the regression algorithm can be used for operational "stitching" of individual frames into a single, continuous image $B = [b(x, j)]$, where, $x = \overline{1, X}$; $j = \overline{1, J}$; $J < K \cdot Y$.

The algorithm is based on the fact that the parameters of the geometric frame alignment functions and are determined based on the coordinates of similar points located in overlapping areas of adjacent images. A similar approach is used in the paper [1] to obtain a panoramic image of the individual frames. Moreover, the parameters of the geometric alignment of the k-th and $(k+1)$-th frames are represented as an affine function:

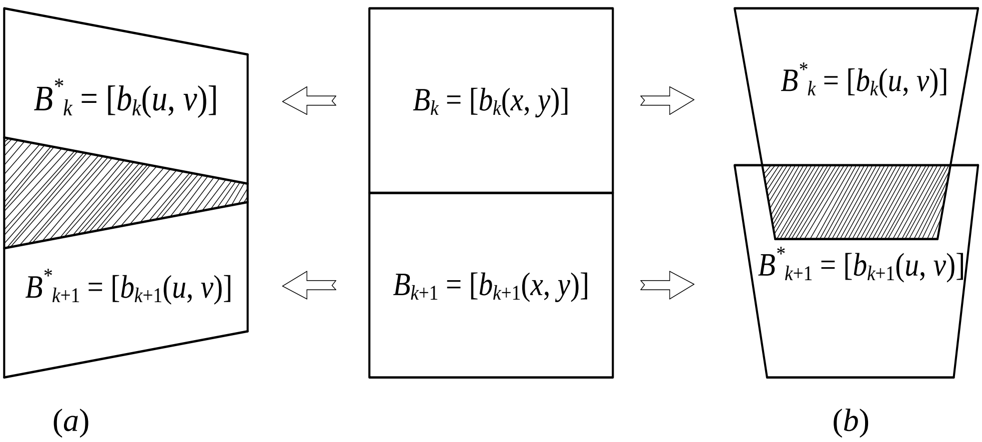
$$x^* = a_0 + a_1 x + a_2 y, \quad y^* = c_0 + c_1 x + c_2 y, \tag{1}$$

where: (x^*, y^*), (x, y) – pixel coordinates of the k-th and $(k+1)$-th frames, respectively; (a_0, a_1, a_2), (c_0, c_1, c_2) – conversion coefficients calculated based on the coordinates of the same points of the k-th and $(k+1)$-th frames. Unfortunately, in the aerial image, each frame is exposed to projective geometric distortions, which are described by equations of the form:

$$u = \frac{a_1 x + a_2 y + a_0}{c_0 x + c_1 y + 1}, \quad v = \frac{d_1 x + d_2 y + d_0}{c_0 x + c_1 y + 1} \tag{2}$$

where: (u, v) – the coordinates of the image point on the earth's surface; (a_0, a_1, a_2), (d_0, d_1, d_2), (c_0, c_1) – transformation parameters, which leads to complicated mutual distortion of "stitchable" frames (Fig.1.). This prevents the use of (1) to form a continuous image B.

Fig. 1. Example of the mutual position of the k-th and (k + 1)-th frames on the earth's surface at the corners of the roll (a) and pitch (b)of the recording device



The objective of the work is to develop mathematical relationships, which is used for the synthesis of the continuous aerial image of the earth's surface from a set of overlapping frames, which are formed in the conditions of an aerial image.

Image stitching algorithm

The formation of a continuous image will be carried out by a sequential geometric transformation of (k+1)-th frame in the coordinate system k using the equation:

$$b_k\left(F_x(x,y),\, F_y(x,y)\right) = b_{k+1}(x,y), \qquad (3)$$

where: $x^* = F_x(x,y)$, $y^* = F_y(x,y)$ – the desired equations of geometric correspondence between the coordinate system of the base $\left(x^*,0,\ y^*\right)$ and the attached (x,0, y) frames.

The features of finding F_x, F_y functions are as follows.

Firstly, due to the differences in the shooting perspective of the k-th and (k+1)-th frames, their common areas are geometrically distorted with respect to each other and generally have different brightness. As a result, the same high-rise objects have a different view in the common areas of matching images. All this makes it difficult to directly use the method of correlation-extreme identification to find the coordinates of the points of the same name and restore the functions F_x, F_y [2, 3].

Secondly, the parameters of the function of geometric alignment of two frames within the common areas are determined by the coordinates of the points of the same name. In this case, the entire (k+1)-th frame is converted so that the law of coordinate correspondence with (k+2)-th frame is violated. That is, faced with the task of constructing such a model coordinate compliance staff, which would extend only to the area of combining frames and does not lead to a change in the geometric dimensions of the synthesized continuous shooting route.

According to this features proposed multistage algorithm for finding the required functions F_x, F_y.

Step 1. Approximate definition of the overlapping area of adjacent frames

To ensure the independence of the correlation function from the brightness differences of adjacent frames, the transition from the brightness values of each pixel to the value of the brightness gradient at a given point is carried out:

$$\begin{aligned} S : B_k &\to G_k = \{g_k(x,y)\}, \\ S : B_{k+1} &\to G_{k+1} = \{g_{k+1}(x,y)\}, \end{aligned} \qquad (4)$$

where: S – Sobel operator [3]; G_k, G_{k+1} – contour representation k-th and (k+1)-th frame.

We assume that the earth's surface shooting is performed with overlapping neighboring frames is not more than 50%. Accordingly, the upper portion of the image G_{k+1} is divided into many square fragments Z_t, $t = \overline{1, T}$. The size of each fragment is chosen so that it is half of the intended area of overlap frames. Denote by dx_t, dy_t fragment offset center Z_t, identified in the determining frame G_k. Offset dx_t, dy_t found from the condition:

$$(dx_t, dy_t) = \arg\min\left[\frac{1}{|Z_t|}\sum_{y=0}^{Y/4}\sum_{x=0}^{X/4}\left|g_{k+1}(x,y) - g_k(x',y')\right|\right], \qquad (5)$$
$$x' = x + dx_t,\ y' = y + dy_t,\ dx_t = dy_t = \overline{0, Y/4},$$

where: $|Z_t|$ – the number of points in the fragment Z_t.

Let's check the validity of the found offsets with the help of the reverse search algorithm [4, 5]. To do this, we will look for the position of the fragmen Z_t in the definition of the G_{k+1}. If the offset found during the forward and reverse search, do not differ by more than 1, then the values of dx_t, dy_t we will be considered reliable.

By averaging received values after checking offsets dx_t, dy_t, we obtain the offsets dx, dy (k+1)-th frame relative to k-th and thereby determine the approximate values of the width W and height H of overlap region k-th and (k+1)-th frames, respectively.

Step 2. Determining the coordinates of the similar objects

For reliable identification of similar objects located in the overlapping parts of the k-th and (k+1)-th frames, the Harris detector is used [4]. This operator allows to select the image feature points in the vicinity in which the image gradient has two dominant direction. To determine the feature points of k-th frame, perform a sequence of steps.

Using the Gaussian filter with 3×3 window perform smoothing of the frame portion, which has an overlap with the(k+1)-th frame

$$B_k = B_k \otimes S_t, \qquad (6)$$

where: $\otimes$ – the convolution operation with the window

$$S_l = [s_{ij}];\ s_{ij} = \left|\frac{1}{2\pi l} e^{-(i^2+j^2)/2l};\ i = j = \overline{1, l};\ (l = 3).\right.$$

Next, using a Sobel operator to image differentiation B_k:

$$V_{kx} = B_k \otimes \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix},\ V_{ky} = B_k \otimes \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \qquad (7)$$

Then, the obtained contour images are again subjected to the Gaussian filter in the window 11×11:

$$
\begin{aligned}
I_{kx} &= [q_{kx}(x,y)],\ I_{kx} = V_{kx}^2 \otimes S_l,\\
I_{ky} &= [q_{ky}(x,y)],\ I_{ky} = V_{ky}^2 \otimes S_l,\\
I_{kxy} &= [q_{kxy}(x,y)],\ I_{kxy} = V_{kxy}^2 \otimes S_l,
\end{aligned} \tag{8}
$$

Here, the window size is determined depending on the spatial resolution frames.

Form the symmetric matrix derivatives at the point (x, y):

$$
I = \begin{bmatrix} q_{kx}(x,y) & q_{kxy}(x,y) \\ q_{kxy}(x,y) & q_{ky}(x,y) \end{bmatrix}. \tag{9}
$$

The feature points have a significant value of the eigenvalues λ_1, λ_2 of the matrix *I*, so we define the set of coordinates of feature points $b_k(x_\xi, y_\xi)$ for which the value of $R(x, y) = \det I - \theta.\ \mathrm{trace}\ I$ maximum where $\theta = 0.04$ – empirical coefficient.

Finally, using the correlation search device for each point $b_k(x_\xi, y_\xi)$ we find its image in the coordinate system $(k+1)$-th frame. The founded values are denoted as $b_{k+1}(x_\xi^*, y_\xi^*)$.

Step 3. Determining the parameters of the mathematical model frame alignment

The desired equations F_x, F_y are constructed based on projective transformation formula (10). To save the converted coordinate matching and subsequent frames provide invariability $(k+1)$-th frame in the overlap region to $(k+2)$-th through a smooth «attenuation transformation» throughout the frame area. To do this, we introduce the function:

$$
h(y) = \begin{cases} 1, & y < H_k, \\ 0, & y > H_{k+1}, \\ 1 - \dfrac{y - H_k}{H_{k+1} - H_k}, & H_k \le y \le H_{k+1}, \end{cases} \tag{10}
$$

where: H_k – the height of the intersection region of *k*-th and $(k+1)$-th frames, H_{k+1} – height of the intersection region $(k+1)$-th and $(k+2)$-th frame.

Modified law of projective transformation is defined as follows:

$$
\begin{aligned}
F_x(x,y) &= h(y)\frac{a_1x^* + a_2y^* + a_0}{c_0x^* + c_1y^* + 1} + (1 - h(y))x,\\
F_y(x,y) &= h(y)\frac{d_1x^* + d_2y^* + d_0}{c_0x^* + c_1y^* + 1} + (1 - h(y))x.
\end{aligned} \tag{11}
$$

Specific values of the transform coefficients computed using the sets of similar pixels $b_k(x_\xi, y_\xi)$, $b_{k+1}(x_\xi^*, y_\xi^*)$, for which the minimum residual function:

$$f = \left(x_\xi - \frac{a_1 x_\xi^* + a_2 y_\xi^* + a_0}{c_0 x_\xi^* + c_1 y_\xi^* + 1} \right)^2 + \left(y_\xi - \frac{d_1 x_\xi^* + d_2 y_\xi^* + d_0}{c_0 x_\xi^* + c_1 y_\xi^* + 1} \right)^2 \quad (12)$$

The minimization function is performed by the method of Levenberg-Marquard using initial approximation $a_1 = d_2 = 1$, $a_0 = a_2 = d_0 = d_1 = c_0 = c_1 = 0$ [3, 4, 6].

Experimental results

The described algorithm is implemented with a software module in the language Python. Used frames generated using aerial digital camera "Agros" system for the experimental testing algorithm. Resolution 7216×5412 pixels per image, the area of the intersection of adjacent frames is from 25% to 50% of the frame.

The number of detected singular points, depending on the type of terrain is given in *Table1*. To determine the parameters of the functions (12), 5 pairs of uniformly distributed similar points are sufficient, which means that the proposed step-by-step algorithm of correlation-extreme identification guarantees the finding of the required number of points for the construction of geometric correspondence functions [6].

Table 1). *Performance analysis of the proposed algorithm*

Scene	Forest	Building	Field
Number of singular points	17-48	25-69	13-45

The residual function (12) for all the frames do not exceed 5 pixels. Fig.2 is a continuous portion of the image formed from the individual images of the mixed type of terrain. Fig.3 (*a*)(*b*) is an enlarged portion of the image at the junction of adjacent frames.

Fig. 2. Detail of the synthesized panoramic image (*a*) *field* (*b*) *forest* (*c*) *building*

Fig. 3. Parts of the synthesized panoramic image (a) forest and (b) building

(*a*)

(*b*)

2. Conclusion

The developed algorithm for the formation of a continuous image of the earth's surface from a set of overlapping frames showed high efficiency in the processing of the aerial image in the absence of information about the frame size, their mutual brightness and coordinate distortions.

References

[1] Naing Linn Aung, Koldaev V. D. *Method of stitching aerial images frames in video surveillance systems.* Scientific and technical journal "news of universities. ELECTRONICS» Volume 23, No. 2, 2018. - P. 194-202.

[2] Koldaev V. D. *Using graph models for image segmentation.* "Scientific space of Europe". Proceedings of the X International scientific-practical conference «Scientific space of Europe-2014» (07-15 April 2014). - Volume 36: Technical Sciences. - Przemysl: Science and education, 2014. - P. 35-38.

[3] Pratt W. *Digital image processing.* - M.: Mir, 1982. 311 p.

[4] Gonzalez R., Woods R. *Digital image processing*. - M.: Technosphere, 2000. 1072 p.

[5] Zaw Win Htet, Kyaw Zaw Ye, Koldaev V.D., Dorogov V.G., Naing Linn Aung. *The Edges Detection in Images Using the Clustering Algorithm*. Proceedings of the 2017 IEEE Russia Section Young Researchers in Electrical and Electronic Engineering Conference (2017 ElConRus) February 1-3, 2017, Art.no. 7910628, Part II. – St. Petersburg, Moscow, «LETI», «MIET», 2017. – P. 609-612.

[6] Koldaev V. D., Gagarina L. G. *Methodological approaches to contour segmentation of images in automated production systems.* Scientific and technical journal «News of higher educational institutions. Electronics» №4 (108), 2014. - P. 64-72.