HIGH TECHNOLOGY LETTERS | Vol. 30 No. 3 | Sep. 2024 | pp. 231-243

doi:10.3772/j.issn.1006-6748.2024.03.002

Ground target localization of unmanned aerial vehicle based on scene matching①

ZHANG Yan(张 岩)*, CHEN Yukun②*, HUANG He*, TANG Simi**, LI Zhi**

(*Beijing Yansou Technology Co., Ltd., Beijing 100085, P. R. China)

(**The 15th Research Institute of China Electronics Technology Group Corporation, Beijing 100083, P. R. China)

Abstract

In order to improve target localization precision, accuracy, execution efficiency, and application range of the unmanned aerial vehicle (UAV) based on scene matching, a ground target localization method for unmanned aerial vehicle based on scene matching (GTLUAVSM) is proposed. The suggested approach entails completing scene matching through a feature matching algorithm. Then, multi-sensor registration is optimized by robust estimation based on homologous registration. Finally, basemap generation and model solution are utilized to improve basemap correspondence and accomplish aerial image positioning. Theoretical evidence and experimental verification demonstrate that GTLUAVSM can improve localization accuracy, speed, and precision while minimizing reliance on task equipment.

Key words: scene matching, basemap, adjustment, feature registration, random sample consensus (RANSAC), unmanned aerial vehicle (UAV)

0 Introduction

Scene matching is a computer vision technology that determines image areas from corresponding scene areas captured by different sensors, or finds their correspondence[1]. Target localization of unmanned aerial vehicle (UAV) using aerial images (video and photo data) and scene matching techniques is the premise of ground target localization of unmanned aerial vehicle based on scene matching (GTLUAVSM)[2]. This method has many advantages, including less tasks and equipment, while achieving high localization precision and avoidance of external disturbances[3].

The research emphasis of GTLUAVSM lies on multi-sensor registration and a localization strategy. Multi-sensor registration requires high robustness and instantaneous feature detection, feature description and model estimation[4-5]. Localization strategies require a good correspondence between a basemap and a target image and depends less on hardware, and more on fast and accurate coordinate calculation[6]. For such a localization method, domestic and overseas scholars have conducted several studies.

Tian[7] proposed a remote sensing image matching and target localization method based on a local feature search with topological constraints. Firstly, the feature detection component used a multiscale corner detection algorithm. Then, a scale invariant feature transform (SIFT)[8] descriptor was used for the feature description component. Finally, mismatching points were eliminated on the basis of topological constraints to achieve image matching and to acquire longitude and latitude coordinates of important points in the image. The localization precision of this method remains to be improved. Additionally, the speed and robustness of the feature matching technique needs enhancing. Wang et al.[9] proposed a fast target localization method of global image registration. First, a wavelet decomposition filter was designed for aerial images. Then, the SIFT method was applied to angular points for feature detection. A SIFT descriptor was then applied for the feature description component of the process. Finally, random sample consensus (RANSAC)[10] and a least squares method were utilized to optimize the homography matrix, achieve image matching and acquire the longitude and latitude coordinates of the points of interest. This method has high localization precision, but its speed and feature matching accuracy remains to be improved. Zhang et al.[11] came up with UAVs scene matching algorithm based on center surrourd extremasstar (CenSurE-star). The process began with Cen-

① Supported by the National Key R&D Program of China (2022YFF0604502).

② To whom correspondence should be addressed. E-mail: mhly123@sina.cn.

Received on Aug. 3, 2023

SurE-star for feature detection. Then, a fast retina feature detection method was used for the feature description step. Finally, the mismatching points were eliminated with RANSAC to complete image matching and acquire longitude and latitude coordinates of the points of interest. This algorithm can satisfy the vision-aided navigation demands of UAVs with high operation efficiency. However, the robustness of feature-matching operator needs to be improved. Zhang et al. [12] proposed a ground target localization method of UAVs based on position and orientation system (POS) and image matching methods. A SIFT algorithm was applied for the homologous registration of an aerial image. Then, a matching area search method was used to automatically generate a basemap. Next, a SIFT algorithm was used for multi-sensor registration of aerial images and its corresponding basemap. Finally, the 2D geographic coordinates of the aerial image were determined. This method uses image data to automatically generate a corresponding basemap to achieve accurate localization. But, the robustness and speed of the feature detector and descriptor need improvement. It can be noticeed that the execution efficiency of RANSAC should be enhanced since the reviewed matching area methods totally depend on POS equipment.

The algorithms reviewed innovate and improve various localization schemes based on scene matching, but the following tasks still remain: (1) finding a corresponding basemap or improving the low-similarity aerial images produced by most methods; (2) improving the robustness of the feature detector and feature descriptor; (3) improving the robustness and speed of model estimation; (4) improving the correlation and usefulness of multi-sensor registration.

To solve the above problems, GTLUAVSM is proposed. First, the proposed feature-matching algorithm based on this article is used to complete scene matching. Then, a heterologous matching robustness estimation method based on homologous matching is proposed to optimize the heterologous matching results. Finally, a corresponding base map generation and model solving algorithm is proposed to improve the base map correspondence and realize the positioning of any point in aerial photography.

1 GTLUAVSM principle

1.1 Algorithm steps

The steps of the GTLUAVSM method are listed below.

Step 1 Acquire aerial images of known geographic information V_1, and continuously update the information by acquiring subsequent aerial images $V_2, V_3, \cdots, V_n$, where n is the total number of images. The picture M_1 corresponds to the area ($i = 1$) that is extracted from the basemap on the digital satellite map. Then, as shown in Fig. 1, the 2D geographical coordinates of four angular points are (w_1, B_1), (W_1, B_1), (W_1, b_1) and (w_1, b_1), where M_1 is a $C_1 \times R_1$ pixel image.

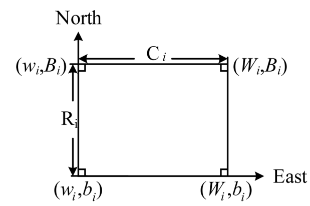


Fig. 1 The structure of basemap

Step 2 First, this study applies camera distortion corrections, atmospheric refraction corrections and earth curvature corrections. These pretreatments are conducted for images V_1, V_2, $\cdots$, V_n. Then, a down-sampling step is carried out for each image $V_1, V_2, \cdots, V_n$. Transverse and longitudinal pixels decrease by χ times (where χ is set by a practical image quality measure). Then, as shown in Fig. 2, an adjacent feature matching (parallel execution) step is carried out. After adjusting a sparse beam method[13], images $A_1, A_2, \cdots, A_{n-1}$ are produced. Then, homography matrices $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_{n-1}$ of the pairwise $V_1, V_2, \cdots, V_n$ are produced by Eq. (1).

$$\boldsymbol{a}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \dfrac{1}{\chi} \end{bmatrix} \boldsymbol{A}_i \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \chi \end{bmatrix} \quad 1 \leqslant i \leqslant n \tag{1}$$

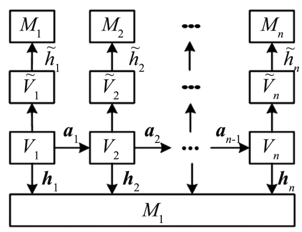


Fig. 2 The structure of scene matching

In Fig. 2, V_i is an aerial image; M_i refers to the base map corresponding to V_i; $\tilde{V}_i$ refers to the image after V_i is transformed according to the homography matrix $\boldsymbol{U}_i$. Finally, the pixel data are adjusted; $\boldsymbol{a}_s$ represents the homography matrix of adjacent images after adjustment; $\tilde{\boldsymbol{h}}_i$ refers to homography matrix from $\tilde{V}_i$ to M_i; $\boldsymbol{h}_i$ refers to a homography matrix from V_i to M_1. The mappable subarea of all aerial images is $c \times r$. n is the total number of images, where $1 \leqslant i \leqslant n$, $1 \leqslant s < n$ and i and s are integers.

Step 3 This study applied an interpolation method to scale the transverse and longitudinal pixel values of V_1 to α_1 and β_1 times. The number of pixels are adjusted to correspond with the transverse and longitudinal pixel count of M_1 to add $\tilde{V}_1$. Then, feature matching is conducted for $\tilde{V}_i$ and M_1 to obtain a homography matrix $\tilde{\boldsymbol{h}}_1$ from $\tilde{V}_i$ to M_1 . Finally, $\boldsymbol{h}_1$ is determined via Eq. (2), where $\alpha_1, \beta_1 > 0$.

$$\boldsymbol{h}_1 = \tilde{\boldsymbol{h}}_1 \begin{bmatrix} \frac{1}{\alpha_1} & 0 & 0 \\ 0 & \frac{1}{\beta_1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Step 2 and Step 3 are executed at the same time. In the two steps, the feature matching process is shown in Fig. 3.

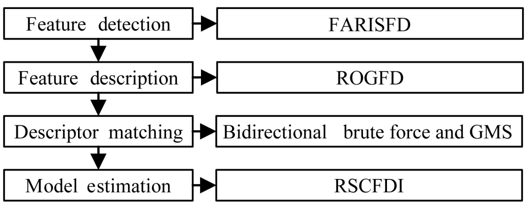


Fig. 3 The feature matching algorithm

Feature matching includes homologous image matching (for aerial images) and multi-sensor image matching (between aerial image and base map). The following order is followed: feature detection, feature description, descriptor matching and model estimation.

It includes following steps. Firstly, keypoints are detected by fast adaptive robust invariant scalable feature detector (FARISFD)[14]. Secondly, characteristics of descriptors is formed by robust overlapped gauge feature descriptor (ROGFD)[15]. Thirdly, matching the descriptors is adopted by bidirectional brute matching based on Euclidean distance and eliminating mismatches is employed by grid-based motion statistics (GMS) for fast, ultra-robust feature correspondence [16]. Finally, homography is worked out by random sample consensus based on feature distance and inliers (RSCFDI)[17].

Step 4 The overall adjustment of multi-sensor registration is conducted to determine $\boldsymbol{H}_{t,t}$. In this step, $\boldsymbol{H}_{t,t}$ is the result after the adjustment for $1 \leqslant t < n$ (where t is an integer and its initial value is 1).

Step 5 First, M_{t+1} is generated according to the generation and model solution method of the corresponding base map. Then, a multi-sensor registration is implemented to find $\tilde{\boldsymbol{h}}_{t+1}$. Finally, $\boldsymbol{h}_{t+1}$ is obtained by performing a number of calculations.

Step 6 Determine whether t is equal to $n-1$. If it is, execute Step 6; if not, change t to $\bar{t} = t + 1$ by replacing t in Step 3 and Step 4 and executing Step 3.

Step 7 $\boldsymbol{H}_{i,n}$ is found via an overall adjustment of multi-sensor registration. $\boldsymbol{H}_{i,n}$ is determined by Eq. (3).

$$\boldsymbol{H}_{i,n} = \begin{bmatrix} H_{11}^{i,n} & H_{12}^{i,n} & H_{13}^{i,n} \\ H_{21}^{i,n} & H_{22}^{i,n} & H_{23}^{i,n} \\ H_{31}^{i,n} & H_{32}^{i,n} & H_{33}^{i,n} \end{bmatrix} \tag{3}$$

Step 8 The 2D geographical coordinates of (X_i, Y_i) for any point (x_i, y_i) in V_i , for the computer image coordinate system, are transformed by Eq. (4). On the digital satellite map, the corresponding elevation information Z_i is found point-wise (X_i, Y_i) to complete target localization.

$$\begin{cases} X_i = w_1 + \dfrac{(W_1 - w_1)(H_{11}^{i,n} x_i + H_{12}^{i,n} y_i + H_{13}^{i,n})}{C_1(H_{31}^{i,n} x_i + H_{32}^{i,n} y_i + H_{33}^{i,n})} \\ Y_i = B_1 + \dfrac{(b_1 - B_1)(H_{21}^{i,n} x_i + H_{22}^{i,n} y_i + H_{23}^{i,n})}{R_1(H_{31}^{i,n} x_i + H_{32}^{i,n} y_i + H_{33}^{i,n})} \end{cases} \tag{4}$$

1.2 Relevant descriptions

(1) GTLUAVSM avoids the use of POS, i. e., the method avoids the error influence resulting from POS system, which weakens the dependence on airborne equipment.

(2) Since feature matching determines a geometrical relationship between homologous and multi-sensor images, this greatly affects target localization precision, accuracy and execution efficiency.

(3) FARISFD and ROGFD enhance robustness of the feature detection and description process. GMS removes mismatched points quickly. RSCFDI improves the execution efficiency substantially, while the RANSAC robustness is guaranteed. Therefore, the feature matching method proposed in this study has strong robustness and speed.

(4) When focal length is small and flight height is large, the slight error of the exterior orientation element of the shooting photocenter in the traditional forward intersection method will lead to large error in the target localization process, while GTLUAVSM has higher localization precision.

2 Robust estimation of multi-sensor registration based on homologous registration

2.1 Algorithm design

(1) To reduce the error among homonymy points, the error of multiple independent observations of a homography matrix should be reduced as much as possible. (2) To avoid extra feature matching steps, the re-

lationship between homography matrices should be leveraged. (3) Since homologous registration is more accurate than multi-sensor registration, homologous registration should be used as much as possible to enhance the accuracy of multi-sensor registration. (4) Observation data should be fully utilized to limit the use of ambiguous observation data and to ensure the adjustment approximates the correct result.

2.2 Algorithm principle

$\boldsymbol{L}$ is set as the homography matrix, as expressed as Eq. (5). The functions $Z_x = (\boldsymbol{L}, x, y)$ and $Z_y = (\boldsymbol{L}, x, y)$ are defined as Eqs (6) and (7).

$$\boldsymbol{L} = \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \tag{5}$$

$$Z_x(\boldsymbol{L}, x, y) = \frac{l_{11}x + l_{12}y + l_{13}}{l_{31}x + l_{32}y + l_{33}} \tag{6}$$

$$Z_y(\boldsymbol{L}, x, y) = \frac{l_{21}x + l_{22}y + l_{23}}{l_{31}x + l_{32}y + l_{33}} \tag{7}$$

Robust estimation of multi-sensor registration based on homologous registration is executed via Steps 1 – 5:

Step 1 Set the weight to $p_m = 1$. Let $\boldsymbol{h}_m$ and a_s be transformed by Eqs (8) and (9) to find $H_{1,f}$ for $1 \leqslant f \leqslant n$, $1 \leqslant m \leqslant f$, $1 \leqslant j, k \leqslant f$, where f, m, j and k are integers.

$$H_{m,f} = \frac{\sum_{j=1}^{f} p_j h^{(j)(m)}}{\sum_{j=1}^{f} p_j} \tag{8}$$

$$h^{(j)(k)} = \begin{cases} h_j \prod_{z=j}^{k-1} a_z^{-1} & j < k \\ h_j & j = k \\ h_j \prod_{z=1}^{j-k} a_{j-z} & j > k \end{cases} \tag{9}$$

Step 2 Calculate d_m and σ by using Eq. (10) and Eq. (11), respectively.

$$d_m = \frac{\int_{x=0}^{c}\int_{y=0}^{r} \sqrt{\left(\begin{array}{c} \left(\begin{array}{c} Z_x(h^{(m)(1)}, x, y) - \\ Z_x(H_{1,f}, x, y) \end{array}\right)^2 + \\ \left(\begin{array}{c} Z_y(h^{(m)(1)}, x, y) - \\ Z_y(H_{1,f}, x, y) \end{array}\right)^2 \end{array}\right)}}{cr} \tag{10}$$

$$\sigma = \sqrt{\frac{\sum_{m=1}^{f} p_m d_m^2}{f - 1}} \tag{11}$$

Step 3 Calculate the equivalent weight $\bar{p}_m$ according to the IGGⅢ scheme[18] using Eq. (12), where $u_m = \dfrac{d_m}{\sigma}$, $k_0 = 1.5$, $k_1 = 3.0$ are eliminated points.

$$\bar{p}_m = \begin{cases} p_m & |u_m| < k_0 \\ \dfrac{p_m k_0}{|u_m|}\left(\dfrac{k_1 - |u_m|}{k_1 - k_0}\right)^2 & k_0 \leqslant |u_m| < k_1 \\ 0 & |u_m| \geqslant k_1 \end{cases} \tag{12}$$

Step 4 Replace p_m with $\bar{p}_m$ in Eq. (12) to update the $H_{1,f}$ approximation.

Step 5 Determine whether the difference between the two estimation results is less than the tolerance. If it is, output $H_{m,f}$; if it is not, go back to Step 2.

2.3 Proof of Eqs (8) and (9)

Proof For the algorithm proposed in this study, the precision of the homologous feature matching is higher than that of multi-sensor feature matching, so a_s is used for adjustment. As shown in Fig. 2, there are f independent observations of the homography matrix that update V_1 until the final M_1 is acquired. The updates are made according to a geometrical relationship between $V_1, V_2, \cdots, V_f$ and M_1: $h_1, h_2 a_1, \cdots, h_f a_{f-1} a_{f-2} \cdots a_1$, where weights of the independent observations are $p_1, p_2, \cdots, p_f$. The correction number for each observation is set to E_m, then:

$$\begin{cases} H_{1,f} = h_1 + E_1 \\ H_{2,f}\, a_1 = h_2\, a_1 + E_2 \\ H_{3,f}\, a_2\, a_1 = h_3\, a_2 a_1 + E_3 \\ \cdots \\ H_{f,f}\, a_{f-1}\, a_{f-2} \cdots a_1 = h_f\, a_{f-1}\, a_{f-2} \cdots a_1 + E_f \end{cases} \tag{13}$$

when $f = 1$, then:

$$H_{1,1} = h_1 \tag{14}$$

when $f > 1$, then the function model of adjustment is

$$\begin{aligned} &\min \sum_{m=1}^{f} (E_m^2(P, Q)) \\ &\text{s.t.} \begin{cases} (h_1(P,Q) + E_1(P,Q)) - \\ ((h_2\, a_1)(P,Q) + E_2(P,Q)) = 0 \\ (h_1(P,Q) + E_1(P,Q)) - \\ ((h_3\, a_2\, a_1)(P,Q) + E_3(P,Q)) = 0 \\ \cdots \\ (h_1(P,Q) + E_1(P,Q)) - \\ ((h_f\, a_{f-1}\, a_{f-2} \cdots a_1)(P,Q) + E_f(P,Q)) = 0 \end{cases} \end{aligned} \tag{15}$$

where $E_m(P, Q)$ is the element of E_m in the Pth row and Qth column, and $1 \leqslant P \leqslant 3$, $1 \leqslant Q \leqslant 3$.

Then, the model function can be solved.

Make:

$$\boldsymbol{P} = \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & P_f \end{bmatrix} \tag{16}$$

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \tag{17}$$

$$\boldsymbol{V} = \begin{bmatrix} E_1(P,Q) \\ E_2(P,Q) \\ \vdots \\ E_f(P,Q) \end{bmatrix} \tag{18}$$

$$\boldsymbol{W} = \begin{bmatrix} (h_1 - h_2 a_1)(P,Q) \\ (h_1 - h_3 a_2 a_1)(P,Q) \\ \vdots \\ (h_1 - h_f a_{f-1} a_{f-2} \cdots a_1)(P,Q) \end{bmatrix} \tag{19}$$

then,

$$\boldsymbol{AV} + \boldsymbol{W} = \boldsymbol{O} \tag{20}$$

where $\boldsymbol{O}$ is a null matrix. According to Lagrange multiplier method, which aims to find the conditional extremum, set K and form the function $\boldsymbol{\Phi}$. Then:

$$\boldsymbol{\Phi} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{PV} - 2\boldsymbol{K}^{\mathrm{T}}(\boldsymbol{AV} + \boldsymbol{W}) \tag{21}$$

Find the first-order derivative for $\boldsymbol{V}$ through $\boldsymbol{\Phi}$, and set it to 0. Then:

$$\frac{\mathrm{d}\boldsymbol{\Phi}}{\mathrm{d}\boldsymbol{V}} = 2\boldsymbol{V}^{\mathrm{T}}\boldsymbol{P} - 2\boldsymbol{K}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{O} \tag{22}$$

The solution is

$$\begin{cases} E_1 = \dfrac{1}{\sum_{j=1}^{f} p_j}\begin{pmatrix} (\sum_{j=1}^{f} p_j - p_1)h_1 + p_2 h_2 a_1 \\ + \cdots + p_f h_f a_{f-1} a_{f-2} \cdots a_1 \end{pmatrix} \\ E_2 = \dfrac{1}{\sum_{j=1}^{f} p_j}\begin{pmatrix} p_1 h_1 + (\sum_{j=1}^{f} p_j - p_2)h_2 a_1 \\ + \cdots + p_f h_f a_{f-1} a_{f-2} \cdots a_1 \end{pmatrix} \\ \vdots \\ E_f = \dfrac{1}{\sum_{j=1}^{f} p_j}\begin{pmatrix} p_1 h_1 + \cdots + \\ (\sum_{j=1}^{f} p_j - p_f) \\ \times h_2 a_1 h_f a_{f-1} a_{f-2} \cdots a_1 \end{pmatrix} \end{cases} \tag{23}$$

Then, combining Eqs (13), (14) and (23). Eq. (8) and Eq. (9) are established.

Similarly, the model function established according to the observations of the homography matrix from V_2, $V_3, \cdots, V_f$ to M_1 can prove that Eq. (8) and Eq. (9) are established.

The proof process is complete.

3 Basemap generation and model solution

3.1 Algorithm design

(1) Since feature matching involves large calculations for each step, then the number of feature matching steps should be reduced as much as possible. (2) An increase in image size will lead to an increase in the number of calculations performed in the feature matching step; therefore, the basemap should be as small as possible. Meanwhile, the basemap should accurately cover the aerial photography area as far as possible. (3) Due to the large difficulty of multi-sensor registration, the differences between aerial images and basemaps should decrease as much as possible. (4) To simplify the adjustment and coordinate calculations, multi-sensor registration of aerial images and their corresponding basemaps should be achieved as far as possible from each other to transform the multi-sensor registration of the aerial images and the same basemap.

3.2 Algorithm principle

The functions $\boldsymbol{Z}_{x\min}(\boldsymbol{L})$, $\boldsymbol{Z}_{x\max}(\boldsymbol{L})$, $\boldsymbol{Z}_{y\min}(\boldsymbol{L})$, $\boldsymbol{Z}_{y\max}(\boldsymbol{L})$ and $\boldsymbol{Z}(\boldsymbol{L})$ are defined below.

$$\boldsymbol{Z}_{x\max}(\boldsymbol{L}) = \max\begin{pmatrix} Z_x(\boldsymbol{L},0,0), \\ Z_x(\boldsymbol{L},c,0), \\ Z_x(\boldsymbol{L},c,r), \\ Z_x(\boldsymbol{L},0,r) \end{pmatrix} \tag{24}$$

$$\boldsymbol{Z}_{y\max}(\boldsymbol{L}) = \max\begin{pmatrix} Z_y(\boldsymbol{L},0,0), \\ Z_y(\boldsymbol{L},c,0), \\ Z_y(\boldsymbol{L},c,r), \\ Z_y(\boldsymbol{L},0,r) \end{pmatrix} \tag{25}$$

$$\boldsymbol{Z}_{x\min}(\boldsymbol{L}) = \min\begin{pmatrix} Z_x(\boldsymbol{L},0,0), \\ Z_x(\boldsymbol{L},c,0), \\ Z_x(\boldsymbol{L},c,r), \\ Z_x(\boldsymbol{L},0,r) \end{pmatrix} \tag{26}$$

$$\boldsymbol{Z}_{y\min}(\boldsymbol{L}) = \min\begin{pmatrix} Z_y(\boldsymbol{L},0,0), \\ Z_y(\boldsymbol{L},c,0), \\ Z_y(\boldsymbol{L},c,r), \\ Z_y(\boldsymbol{L},0,r) \end{pmatrix} \tag{27}$$

$$\boldsymbol{Z}(\boldsymbol{L}) = \begin{bmatrix} 1 & 0 & -\boldsymbol{Z}_{x\min}(\boldsymbol{L}) \\ 0 & 1 & -\boldsymbol{Z}_{y\min}(\boldsymbol{L}) \\ 0 & 0 & 1 \end{bmatrix} \tag{28}$$

The basemap generation and model determination are executed as Steps 1 - 3:

Step 1 Calculate the 2D geographical coordinates of the four angular points of the corresponding basemap M_μ of V_μ for the points: (w_μ, B_μ), (W_μ, B_μ), (W_μ, b_μ) and (w_μ, b_μ). The characteristics are as follows.

$$\begin{cases} w_\mu = w_1 + \dfrac{(W_1 - w_1)\boldsymbol{Z}_{x\min}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}^{-1})}{C_1} \\ W_\mu = w_1 + \dfrac{(W_1 - w_1)\boldsymbol{Z}_{x\max}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}^{-1})}{C_1} \\ B_\mu = B_1 + \dfrac{(b_1 - B_1)\boldsymbol{Z}_{y\min}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}^{-1})}{R_1} \\ b_\mu = B_1 + \dfrac{(b_1 - B_1)\boldsymbol{Z}_{y\max}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}^{-1})}{R_1} \\ 1 < \mu \leqslant n \end{cases} \tag{29}$$

Extract the area pictures from the digital satellite map and set it to M_μ ; then, transform V_μ using the homography matrix $\boldsymbol{U}_\mu$ to find $\widehat{V}_\mu$. $\boldsymbol{U}_\mu$ is shown as

$$\boldsymbol{U}_\mu = \boldsymbol{Z}(H_{\mu-1,\mu-1}\, a_{\mu-1}^{-1})\boldsymbol{H}_{\mu-1,\mu-1}\, \boldsymbol{a}_{\mu-1}^{-1} \tag{30}$$

Step 2 Interpolation method is used to scale transverse and longitudinal pixel numbers of $\widehat{V}_\mu$ to α_μ and β_μ times, and they are adjusted to be same as transverse and longitudinal pixel numbers of M_μ to find $\tilde{V}_\mu$. Then, the feature matching step is conducted for $\tilde{V}_\mu$ and M_μ to obtain the homography matrix $\tilde{\boldsymbol{h}}_\mu$ from $\tilde{V}_\mu$ to M_μ .

Step 3 Calculate h_μ as below.

$$\boldsymbol{h}_\mu = [\boldsymbol{Z}(\boldsymbol{H}_{\mu-1,\mu-1}\, \boldsymbol{a}_{\mu-1}^{-1})]^{-1}\, \boldsymbol{g}_\mu U \tag{31}$$

$$\boldsymbol{g}_\mu = \begin{cases} \tilde{\boldsymbol{h}}_\mu \boldsymbol{Z}_\mu & e_\mu \leqslant k\, E_\mu \\ \boldsymbol{Z}_\mu & e_\mu > k\, E_\mu \end{cases} \tag{32}$$

$$\begin{cases} e_\mu = \sqrt{\begin{array}{l}(\boldsymbol{Z}_{x\max}(\tilde{h}_\mu) - \boldsymbol{Z}_{x\min}(\tilde{h}_\mu))^2 \\ + (\boldsymbol{Z}_{y\max}(\tilde{h}_\mu) - \boldsymbol{Z}_{y\min}(\tilde{h}_\mu))^2\end{array}} \\ E_\mu = \sqrt{c^2 + r^2} \end{cases} \tag{33}$$

$$\boldsymbol{Z}_\mu = \begin{bmatrix} \dfrac{1}{\alpha_\mu} & 0 & 0 \\ 0 & \dfrac{1}{\beta_\mu} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{34}$$

3.3 Relevant descriptions

(1) Multi-sensor registration after V_μ is transformed to $\tilde{V}_\mu$. This is done to reduce the differences between aerial images and the base map.

(2) The implementation of Eq. (1) and Eq. (2) is the same as that of Eq. (32), that is, calculate a homography matrix of an original image transformation by changing the image resolution and employing the geometrical relationship. The matching algorithm based on feature points has excessive image detail and will not change the transformation matrix by much. The image size will simply reduce the operation efficiency of the feature matching step. In this method, the resolution ratio of images is reduced in the feature matching step and then a homography matrix of the original image transformation is determined by Eq. (1), which ensures the texture detail of the matched images and improves matching precision. Any large differences in the image detail of the same scene will result in a large multi-sensor registration scale transformation, thus increasing the difficulty of multi-sensor registration. In this method, a multi-sensor image size is adjusted to approximate the resolution ratio. Then, a homography matrix of the original image transformation is determined according to Eq. (2) and Eq. (32). This reduces the difficulty of the multi-sensor registration step.

(3) If $\boldsymbol{Z}_{x\min}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}^{-1}) < 0$ or $\boldsymbol{Z}_{y\min}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}^{-1}) < 0$, the image texture loss will be caused after the transformation, which may lead to inaccurate feature matching. If $\boldsymbol{Z}_{x\min}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}^{-1}) > 0$ or $\boldsymbol{Z}_{y\min}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}^{-1}) > 0$, the image size will increase after transformation and this will increase the number of calculations of feature matching step and waste memory. Thus, the function $\boldsymbol{Z}(\boldsymbol{L})$ can ensure the image area after the transformation, located at the image center, is applied, which fixes the map size to the largest extent to avoid texture loss, while decreasing the number of calculations and the memory expenditure.

(4) Eq. (32) and Eq. (33) aim to prevent large-area inaccuracies of image transformations caused by multi-sensor mismatching. Under the condition of the correct multi-sensor registration, e_μ and E_μ differ by a small amount. Under the condition of wrong multi-sensor registration, e_μ and E_μ differ by a large amount. So, in this experiment, the value of k is 5, and this value only has a little influence on the algorithm. In practical applications, the parameters can be adjusted by the operation.

3.4 Proof of Eqs (4), (29) and (31)

Proof According to the computer image coordinate system and the definition of a homography matrix, where $\bar{\boldsymbol{K}}_i$ and $\boldsymbol{K}_\mu$ are nonzero constants, and they can be solved via Eq. (35) and Eq. (36); (p,q) refers to the point of M_1 under computer image coordinate system.

$$\begin{bmatrix} p \\ q \\ 1 \end{bmatrix} = \bar{\boldsymbol{K}}_i\, \boldsymbol{H}_{i,f} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{35}$$

$$\begin{bmatrix} x_\mu \\ y_\mu \\ 1 \end{bmatrix} = \boldsymbol{K}_{\mu-1}\, \boldsymbol{a}_{\mu-1} \begin{bmatrix} x_{\mu-1} \\ y_{\mu-1} \\ 1 \end{bmatrix} \tag{36}$$

Combining Eps (35) and (36),

$$\begin{bmatrix} p \\ q \\ 1 \end{bmatrix} = \bar{\boldsymbol{K}}_{\mu-1} \boldsymbol{K}_{\mu-1} \boldsymbol{H}_{\mu-1,f} \boldsymbol{a}_{\mu-1}^{-1} \begin{bmatrix} x_{\mu-1} \\ y_{\mu-1} \\ 1 \end{bmatrix} \tag{37}$$

Let $f = \mu - 1$ and introduce the points $(0,0)$, $(c,0)$, (c,r), and $(0,r)$ into Eq. (37) to obtain image coordinates of four points in M_1. The image coordinates of four angular points of the bounding rectangle formed by the four points are calculated via the expression below:

$$(\boldsymbol{Z}_{x\min}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}), \boldsymbol{Z}_{y\min}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}))$$
$$(\boldsymbol{Z}_{x\max}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}), \boldsymbol{Z}_{y\min}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}))$$
$$(\boldsymbol{Z}_{x\max}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}), \boldsymbol{Z}_{y\max}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}))$$
$$(\boldsymbol{Z}_{x\min}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}), \boldsymbol{Z}_{y\max}(\boldsymbol{H}_{\mu-1,\mu-1}\boldsymbol{a}_{\mu-1}))$$

Since the basemap is a part of the digital satellite map, it is used as a map projection. Then, a geometrical relationship is found and shown in Fig. 1. Then:

$$\begin{cases} \dfrac{X_i - w_1}{W_1 - w_1} = \dfrac{p}{C_1} \\ \dfrac{Y_i - B_1}{b_1 - B_1} = \dfrac{q}{R_1} \end{cases} \tag{38}$$

So, Eq. (29) is established.

Combining Eq. (35) and Eq. (38) and letting $f = n$, Eq. (4) can be determined.

Similarly, Eq. (31) is found by using the geometrical relationship between V_i and the geometrical relationship between V_i and M_i.

The proof process is finished.

3.5 Proof of Eqs (1), (2) and (32)

Proof The homography matrix from image a to image b is set to $\boldsymbol{h}_{ab}$. An interpolation method is used to scale the transverse and longitudinal pixel count from a to α and β times to find image A. The transverse and longitudinal pixel numbers of b are scaled to κ and ω times to find image B, for $\alpha, \beta, \kappa, \omega > 0$.

Under the computer image coordinate system, the homonymy point of any point (x_a, y_a) in a is (x_A, y_A) in A. The homonymy point of any point (x_b, y_b) in b is (x_B, y_B) in B. Then:

$$\begin{cases} X_a = \alpha x_a \\ Y_a = \beta y_a \\ X_b = \kappa x_b \\ Y_b = \omega y_b \end{cases} \tag{39}$$

In accordance with the computer image coordinate system and the definition of homography matrix, the homonymy point (x_{ab}, y_{ab}) in b of any point (x_a, y_a) in a is shown in Eq. (40), where k_1 is a nonzero constant, which can be solved by Eq. (40).

$$\begin{bmatrix} x_{ab} \\ y_{ab} \\ 1 \end{bmatrix} = k_1 \boldsymbol{h}_{ab} \begin{bmatrix} x_a \\ y_a \\ 1 \end{bmatrix} \tag{40}$$

Combine Eq. (39) and Eq. (40), the homonymy point (x_{Ab}, y_{Ab}) in b of any point (x_A, y_A) in A is shown in Eq. (41), where k_2 is a nonzero constant, which can be solved using Eq. (41).

$$\begin{bmatrix} x_{Ab} \\ y_{Ab} \\ 1 \end{bmatrix} = k_2 \boldsymbol{h}_{ab} \begin{bmatrix} \frac{1}{\alpha} & 0 & 0 \\ 0 & \frac{1}{\beta} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_A \\ y_A \\ 1 \end{bmatrix} \tag{41}$$

Similarly, the homonymy point in B of any point (x_A, y_A) in A is expressed in Eq. (42), where k_3 is a nonzero constant, which can be solved using Eq. (42).

$$\begin{bmatrix} \frac{1}{\kappa} & 0 & 0 \\ 0 & \frac{1}{\omega} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{AB} \\ y_{AB} \\ 1 \end{bmatrix} = k_3 \boldsymbol{h}_{ab} \begin{bmatrix} \frac{1}{\alpha} & 0 & 0 \\ 0 & \frac{1}{\beta} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_A \\ y_B \\ 1 \end{bmatrix} \tag{42}$$

The homography matrix $\boldsymbol{H}_{ab}$ from image A to image B is shown in Eq. (43).

$$\boldsymbol{H}_{ab} = \begin{bmatrix} \kappa & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{h}_{ab} \begin{bmatrix} \frac{1}{\alpha} & 0 & 0 \\ 0 & \frac{1}{\beta} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{43}$$

Let $\kappa, \omega, \alpha, \beta = \chi$ and Eq. (1) is established. Let $\kappa, \omega = 1$; $\alpha = \alpha_1$; $\beta = \beta_1$; and Eq. (2) is established. Let $\kappa, \omega = 1$; $\alpha = \alpha_i$; $\beta = \beta_i$; $1 < i \leqslant n$; and Eq. (32) is established.

The proof process is finished.

4 Experimental verification

4.1 Experimental setup

4.1.1 Experimental platform parameters

Laptop configuration: the central processing unit (CPU) is a 4th generation 2.5 GHz i7 system with a 64 bit Win10, programming environment and a Visual Studio 2015 linked with OpenCV 3.00.

4.1.2 Dataset

A region is chosen for the experiment. The technical parameters are shown in Table 1. The UAV video frames and relevant data from a Google digital satellite map are shown in Fig. 4 and listed in Table 2. The shooting time difference between the video frames and the satellite images is 18 months, and the difference in the image resolution ratio is 2. Light conditions differ greatly. The rotation angle of some images exceed 180 °. The change in the points of sight is large, and there is motion fuzziness and added noise influence.

Table 1 The main technical parameters used for experiments

Name	Parameters
Aerial photo time	August, 2016
Aerial camera device	DJI S800 six-rotor drone
Aerial camera	SONY, Zenmuse PTZ system
Aerial camera fixed focus /mm	2.130
Video size /pixels	1 920 × 1 080
Pixel size /mm	0.009
Field of view /degree	35.57 × 26.83
Maximum distortion /mm	0.015
Area range /km	3
Maximum terrain relief /m	373
GPS data update rate /s	1
GPS data initialization /min	3
GPS static observation /min	3
GPS eccentric component /m	2.049, −0.501, 1.381
IMU eccentric component /m	0.000, −0.201, 0.427

(a) The partial unmanned aerial vehicle video frames

(b) The partial Google digital satellite map

Fig. 4 The partial video frames and digital satellite map used for experiments

Table 2 The POS data of the partial unmanned aerial vehicle video frames used for experiments

Image	1	2	…
Latitude /°	34.59	34.59	…
Longitude /°	110.13	110.12	…
Altitude /m	2 302.28	2 306.48	…
Roll /°	−2.03	0.13	…
Pitch /°	0.32	0.18	…
Heading /°	258.34	257.02	…

4.1.3 Experimental objects and related parameter settings

Object 1 The OpenCV parameter settings of the more representative detectors and descriptors in recent years are as follows:

SIFT: feature detector: group number is 4, layer number is 4, contrast threshold is 0.04, edge threshold is 10.00; feature descriptor is 128 dimensions.

Speeded up robust features (SURF)[19]: feature detector: fast-Hessain is 0.6, group number is 4, layer number is 4, non rotation-invariant; feature descriptor is 64 dimensions.

Binary robust invariant scalable keypoints (BRISK)[20]: feature detector: threshold is 30, group number is 4; feature descriptor: pattern scale is 1.

KAZE[21]: feature detector: group number is 4, layer number is 4, diffusivity type is DIFF_CHARBONNIER, threshold is 0.001, non rotation-invariant; feature descriptor is M-SURF of 128 dimensions.

Accelerated-KAZE[22]: feature detector: group number is 4, layer number is 4, diffusivity type is DIFF_CHARBONNIER, threshold is 0.001, non rotation-invariant; feature descriptor is M-SURF of 128 dimensions.

To make sure the number of groups and layers do not influence the objectivity of the experiment, the number of groups and layers of all traditional detectors is 4, and the other parameters are set by default per the OpenCV software.

Object 2 Different sub-methods of the target localization method are shown in Table 3. And the steps and parameters of feature matching algorithm are shown in Table 4.

Table 3 The numbers of different submethods

Submethods	A	B
1	Accelerated-KAZE feature matching	Feature matching in this article
2	None	Robust estimation method for heterologous matching based on homologous matching
3	Matching area search method	Corresponding basemap generation and model solving algorithm
4	Ref. [23] method (using this method to do double slice forward intersection)	None

Table 4 The steps and parameters of feature matching algorithm

Steps	AKAZE	The proposed feature matching algorithm
Feature detection	Accelerated-KAZE (OpenCV default parameters)	FARISFD (default parameters)
Feature description	Accelerated-KAZE (OpenCV default parameters)	ROGFD (default parameters)
Descriptor matching	Brute matching based on Euclidean distance (OpenCV default parameters)	Brute matching based on Euclidean distance and GMS (OpenCV default parameters)
Model estimation	RANSAC (OpenCV default parameters)	RSCFDI (threshold: 60,100)

4.1.4 Experimental objects

To measure the robustness and operation efficiency of the algorithm and compare it with other algorithms, four indexes are used for evaluation purposes: precision mean square error, average matching time-consuming, localization mean square error and average time-consuming per frame.

Index 1 The measurement index of matching precision is precision mean square error. The definition is shown in Eq. (44) and Eq. (45), where Z is the precision mean square error; m is the total number of matches; R_i is the precision ratio of the ith match; C_i is the number of correct matching pairs found by the algorithm in the ith match; P_i is the number of all matching pairs found by the algorithm in the ith match.

$$Z = \frac{1}{m}\sum_{i=1}^{m}(R_i - 1)^2 \tag{44}$$

$$R_i = \frac{C_i}{P_i} \tag{45}$$

Index 2 The measurement index of the operation efficiency of a multi-sensor registration algorithm is the average matching time-consuming. The characterization of this process is shown in Eq. (46), where H is the average matching time-consuming; L_i is the matching time of the ith; m is the total number of matches.

$$H = \frac{1}{m}\sum_{i=1}^{m} L_i \tag{46}$$

Index 3 The measurement index of the localization precision is the localization mean square error. The characterization of the process is shown in Eq. (47) and Eq. (48), where M is the localization mean square error; S_i is the localization error of the kth object; n is the total number of objects; (X_k, Y_k, Z_k) is the observed value of the 3D-coordinate of the kth object; (X_k, Y_k, Z_k) is the theoretical value of the 3D coordinate of the kth object.

$$M = \frac{1}{n}\sum_{k=1}^{n} S_k^2 \tag{47}$$

$$S_k = \sqrt{(X_k - x_k)^2 + (Y_k - y_k)^2 + (Z_k - z_k)^2} \tag{48}$$

Index 4 The measurement index of operation efficiency of the localization algorithm is the average time-consuming per frame. The characterization of the process is shown in Eq. (49), where T is the average time-consuming per frame; t is the total time consumed by the localization algorithm; N is the total number of frames.

$$T = \frac{t}{N} \tag{49}$$

4.1.5 Experimental process

Experiment 1 First, 40 video frames of different flight strips and satellite images in the corresponding area are chosen. A multi-sensor homography matrix is calculated via a manual point selection. Based on the criterion of variable control, SIFT, SURF, KAZE, Accelerated-KAZE, BRISK and FARISFD + ROGFD are used for matching via the multi-sensor registration process shown in Fig. 5. Then, the mean square error of the precision ratio of each algorithm and the average matching time of each algorithm are calculated.

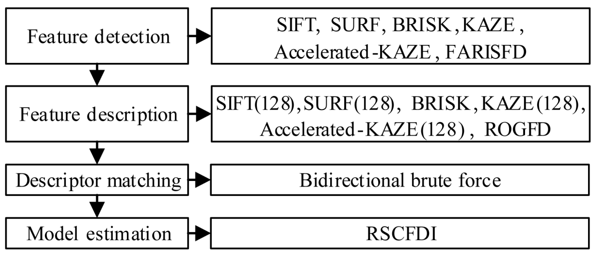


Fig. 5 The structure of feature matching

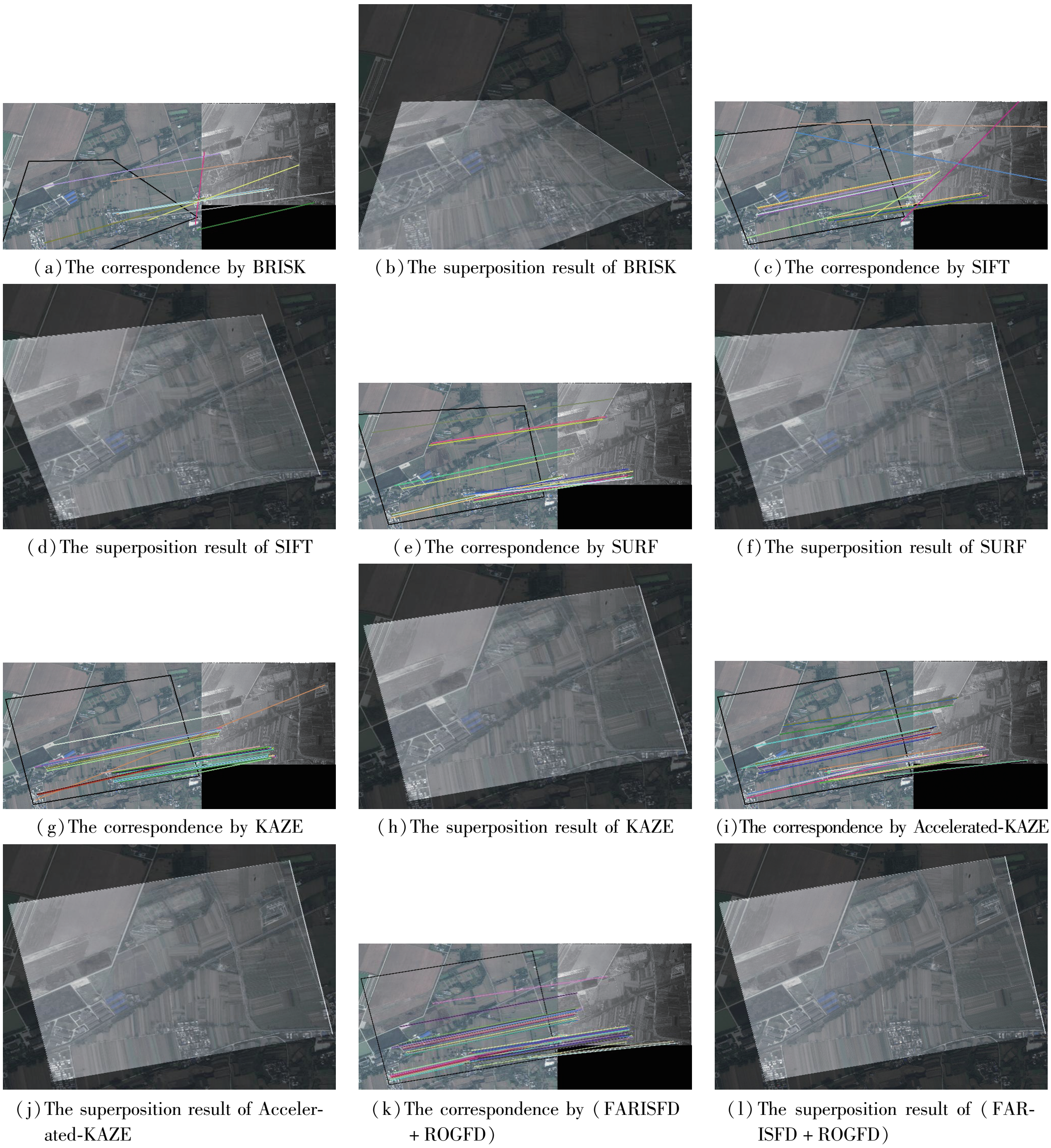
As shown in Fig. 5, multi-sensor registration is conducted in the following order: pretreatment, feature detection, feature description, descriptor matching and model estimation. First, the feature point is found by the detector. Second, the descriptor is utilized to generate the description of the feature point. Third, a violent matching method is used for bilateral matching of the descriptor. Finally, a homography matrix is obtained via a RANSAC calculation.

Experiment 2 First, 200 aerial images of different flight strips have been chosen. A1 + A2 + A3, B1 + A2 + A3, B1 + A2 + B3, B1 + A2 + A4 and B1 + B2 + B3 (GTLUAVSM) are used for localization based on the criterion of variable control. Then, 10 measurement points are selected at random from each aerial image. There are 2 000 measurement points in total. The geographic information from the Google digital satellite map is used as a theoretical value to find the mean square error of localization for each algorithm. Every time per image is calculated for each algorithm.

4.2 Experimental results and analyses

4.2.1 Results and analyses of Experiment 1

To clearly and visually compare and analyze the experimental results, a color line is used to express the connecting line of the homonymy points. Then, the black box is applied to express the perspective transformation result of the UAV video frame according to the matching relation. At the same time, the effect diagram of a map layer overlapping is found. For a frame, the matching results of the different algorithms are shown in Fig. 6. The precision ratio curves of different algorithms are shown in Fig. 7. The matching data statistics are shown in Table 5.



(a) The correspondence by BRISK (b) The superposition result of BRISK (c) The correspondence by SIFT (d) The superposition result of SIFT (e) The correspondence by SURF (f) The superposition result of SURF (g) The correspondence by KAZE (h) The superposition result of KAZE (i) The correspondence by Accelerated-KAZE (j) The superposition result of Accelerated-KAZE (k) The correspondence by (FARISFD + ROGFD) (l) The superposition result of (FARISFD + ROGFD)

Fig. 6 The results of registration

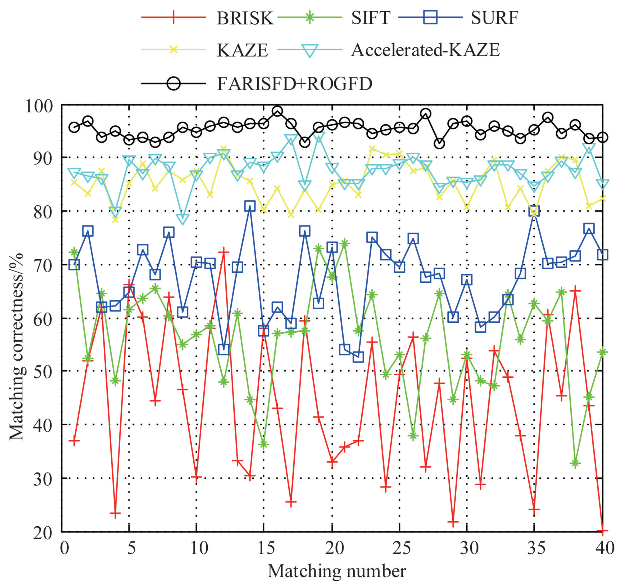


Fig. 7 The accuracy curves

Table 5 The data statistics of registration

Method	Precision mean square error	Average matching time-consuming /s
BRISK	3 269.92	1.42
SIFT	2 007.63	9.11
SURF	1 108.56	4.61
KAZE	230.82	17.37
Accelerated-KAZE	164.61	3.46
FARISFD + ROGFD	23.83	2.78

The experimental results are analyzed as follows.

(1) BRISK is not applicable to multi-sensor registration that has been simplified, has an approximate detection structure and a binary description structure with large differences in illuminance, fuzziness degree and rotation angle. Its matching precision is low, but the speed is very fast.

(2) The matching precision of KAZE is high. This is appropriate for a point of sight, rotation and scale transformation; it shows good robustness, which benefits from the design of the Hessian local maximum point and overlapping strip after different scales in each layer are normalized in the scale space. But, the speed is poor.

(3) Both the matching precision and speed of Accelerated-KAZE are higher than those of KAZE, because the performance of fast nonlinear scale space is stronger than nonlinear scale space.

(4) SURF is applicable to multi-sensor registration, and the speed improves greatly. This method completely meets the speed requirement, and guarantees the needed matching precision, which benefits from the application of a fast Hessian matrix. Its drawback is that it is sensitive to fuzzy transformations and illuminance transformations.

(5) Although SIFT is inferior to SURF on the whole, it shows good rotation robustness when illumination and fuzziness change largely, which benefits from the removal of any marginal responses and gradient histogram statistics. But the trade-off is that the speed is not efficient.

(6) Both the matching precision and speed of FARISFD + ROGFD are higher than those of KAZE. The reasons are as below. First, FARISFD adaptively selects the number of scale space groups, and the scale space based on a transition layer is constructed. Then, a feature score is calculated on the basis of AGAST. Finally, traditional correction methods at the sub-pixel level is simplified, which enhances the robustness and speed of the feature detection step. ROGFD uses a Scharr operator to calculate the image gradient. Then, a second-order standard partial derivative is found to construct an overlapping descriptive grid. In the end, weighting, summation and normalization of the neighborhood responses are conducted, which enhances the robustness and the speed of the feature description step. Thus, the matching effect is good.

4.2.2 Results and analyses of Experiment 2

To compare and analyze experimental results clearly and visually, all pixels in the images are mapped to the Google digital satellite map given the localization results. Some localization results are shown in Fig. 8. A localization error curve for all measurement points is shown in Fig. 9. The localization data statistics is listed in Table 6.

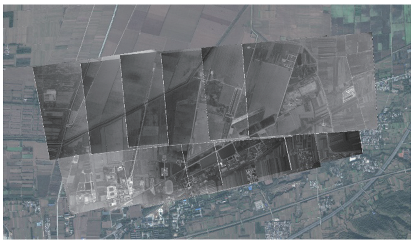
Table 6 The localization data statistics

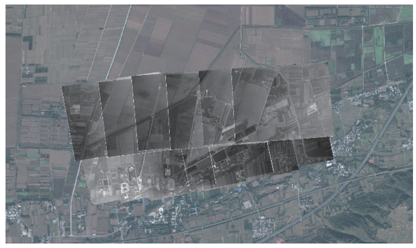
Methods	Localization mean square error	Average time-consuming per frame /s
A1 + A2 + A3	3 191.24	8.11
B1 + A2 + A3	578.61	6.61
B1 + A2 + B3	315.61	4.51
B1 + A2 + A4	13 617.25	3.23
B1 + B2 + B3	225.24	4.56

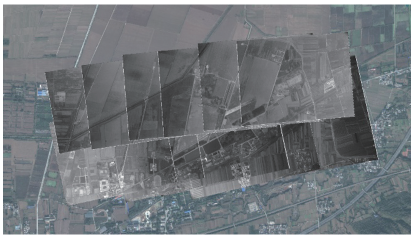
The experimental results are analyzed as follows.

(1) The five curves present random fluctuation trends with a unit of 10 points. This is because the measurement points are located by a homography matrix between images. Thus, the localization errors of the measurement points in the same image are near each other.


(a) The localization result of (A1 + A2 + A3)


(b) The localization result of (B1 + A2 + A3)


(c) The localization result of (B1 + A2 + B3)


(d) The localization result of (B1 + A2 + A4)

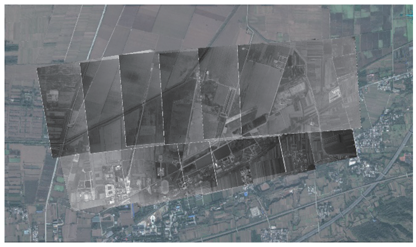

(e) The localization result of (B1 + B2 + B3)

Fig. 8 The localization results

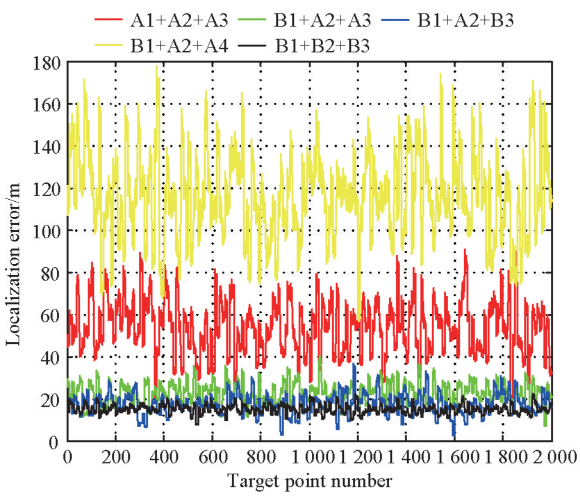


Fig. 9 The localization error curves

(2) The comparison between A1 + A2 + A3 and B1 + A2 + A3 verifies that the proposed feature matching method largely improves precision, accuracy and execution efficiency of target localization based on scene matching.

(3) Compared with the matching area searching method, the basemap generation and model solution enhance the correspondence of the basemap, simplifies the matching area calculations and improves precision, accuracy, and speed of target localization based scene matching. This is accomplished to avoid the use of POS and reduce the dependence on airborne equipment.

(4) Compared with the forward intersection method in Ref. [23], the basemap generation and model solution have higher localization precision.

(5) Multi-sensor registration and the robust estimation algorithm based on homologous registration carries out an overall adjustment for all multi-sensor registration via a robust least square method. Although homologous registration error will reduce the localization precision slightly, the adjustment method lowers the gross error and system error of the overall multi-sensor registration. So, localization precision is enhanced greatly. Relative to homologous and multi-sensor feature matching, the calculation amount of the adjustment algorithm has very little influence on the whole analysis.

5 Conclusions

The feasibility and advantage of GTLUAVSM are verified by theoretical and experimental methods. The conclusions can be listed below.

(1) The feature matching method improves precision, accuracy, and speed of target localization based on scene matching.

(2) Basemap generation and model solution not only improve precision, accuracy and execution efficiency of target localization based on scene matching, but also avoids the error influence of a POS system and reduces the method's dependence on airborne equipment.

(3) The multi-sensor registration robust estimation algorithm based on homologous registration lowers the gross error and system error of multi-sensor registration and improves the localization precision.

(4) Precision, accuracy, and speed of GTLUAVSM are high. Its engineering practice value is also high.

(5) This work recommends the following application scopes and limitations. GTLUAVSM will benefit target localization of small UAVs and UAV-based target localization under interference environments. GTLUAVSM has high requirements for the robustness and speed of feature matching, which largely depends on the basemap.

References

[1] YU Q F, SHANG Y, LIU X C, et al. Full-parameter vision navigation based on scene matching for aircrafts [J]. Science in China Series F: Information Sciences, 2014, 57(5): 1-10.

[2] MARINAKIS Y. An improved particle swarm optimization algorithm for the capacitated location routing problem and for the location routing problem with stochastic demands [J]. Applied Soft Computing, 2015, 37: 680-701.

[3] YANG B S, DONG Z. A shape-based segmentation method for mobile laser scanning point clouds [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2013, 81: 19-30.

[4] HADIZADEH H. Histogram and average neighborhood structure of local lacunarity and fractal features for noise-robust and rotation-invariant texture description [C]// Proceedings of the 9th Iranian Conference on Machine Vision and Image Processing. Tehran, Iran: IEEE, 2015: 9-13.

[5] YING M L, NICHOLAS R G. Predictive RANSAC: effective model fitting and tracking approach under heavy noise and outliers [J]. Computer Vision and Image Understanding, 2017, 161: 93- 113.

[6] CAO S X, JIANG J, ZHANG G J, et al. An edge based scale and affine-invariant algorithm for remote sensing image registration [J]. International Journal of Remote Sensing, 2013, 34(7): 2301-2326.

[7] TIAN Y. Remote sensing image matching and target location based on local features and topological constraints [D]. Boston: Northeastern University, 2014.

[8] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[9] WANG Z Q, CHENG H, YANG G. Fast target location method of global image registration [J]. Infrared and Laser Engineering, 2015, 44(S): 225-229.

[10] FISCHLER M, BOLLES R. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. Communication of the ACM, 1981, 24(6):381-395.

[11] ZHANG W Y, LI Z, WANG Y J. UAV scene matching algorithm based on CenSurE-star feature [J]. Chinese Journal of Scientific Instrument, 2017, 38(2): 462-470. (In Chinese)

[12] ZHANG Y, LI J Z, LI D L, et al. The research of UAV targeting based on POS and image matching [J]. Journal of Ordnance Engineering College, 2015, 27(1): 39-45.

[13] LOURAKIS M I A, ARGYROS A A. SBA: a software package for generic sparse bundle adjustment [J]. ACM Transactions on Mathematical Software, 2009, 36(1): 26-30.

[14] ZHANG Y, LI J Z, LI D L, et al. Fast adaptive robust invariant scalable feature detector [J]. Systems Engineering and Electronics, 2017, 45(6): 1406-1413.

[15] ZHANG Y, LI J Z, LI D L, et al. Robust overlapped gauge feature descriptor[J]. Transactions of Beijing Institute of Technology, 2017(8): 842-847.

[16] BIAN J W, LIN W Y, MATSUSHITA Y, et al. GMS: grid-based motion statistics for fast, ultra-robust feature correspondence [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017: 4181-4190.

[17] ZHANG Y, SUN S Y, HU Y J, et al. Random sample consensus based on feature distance and inliers [J]. Journal of Electronics & Information Technology, 2018, 40(4): 928-935.

[18] YANG Y X. Theory of robust estimation and its application [M]. Beijing: August 1st Publishing House, 1993.

[19] BAY H, TUYTELAARS T, VAN G L. SURF: speeded up robust features [C]//Proceedings of of the 9th European Conference on Computer Vision. Graz, Austria: Springer, 2006: 404-417.

[20] LEUTENEGGER S, CHLI M, SIEGWART R Y. BRISK: binary robust invariant scalable keypoints [C]//Proceedings of the 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 2548-2555.

[21] PABLO F, ADRIEN B, ANDREW J. KAZE features [C] //Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer, 2012: 214-227.

[22] PABLO F, JESUS N, ADRIEN B. Fast explicit diffusion for accelerated features in nonlinear scale spaces [C] // Proceedings of 2013 British Machine Vision Conference. Bristol, UK: BMVC, 2013: 2157-2163.

[23] LI Z M, BIAN S F, QU Y. Robust total least squares estimation of space intersection appropriate for multi-images [J]. Acta Geodaetica et Cartographica Sinica, 2017, 46 (5): 593-604.

ZHANG Yan, born in 1991. He received his B. E. degree from Jilin University in 2009, and his Ph. D degree in Electronic and Optical Engineering from Army Engineering University in 2019. His research interests include computer vision, photogrammetry and information processing.