SCIENCE CHINA
Technological Sciences

• Article •

July 2015 Vol.58 No.7: 1196–1208
doi: 10.1007/s11431-015-5828-x

Vision navigation for aircrafts based on 3D reconstruction from real-time image sequences

ZHU ZunShang[1,2*], SU Ang[1,2], LIU HaiBo[1,2], SHANG Yang[1,2] & YU QiFeng[1,2]

[1] *College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China;*
[2] *Hunan Key Laboratory of Videometrics and Vision Navigation, Changsha 410073, China*

Received November 12, 2014; accepted April 5, 2015; published online June 9, 2015

In this paper, we propose a novel vision navigation method based on three-dimensional (3D) reconstruction from real-time image sequences. It adapts 3D reconstruction and terrain matching to establish the correspondence between image points and 3D space points and the terrain reference (by using a digital elevation map (DEM)). An adaptive weighted orthogonal iterative pose estimation method is employed to calculate the position and attitude angle of the aircraft. Synthesized and real experiments show that the proposed method is capable of providing accurate navigation parameters for a long-endurance flight without using a global positioning system or an inertial navigation system (INS). Moreover, it can be combined with an INS to achieve an improved navigation result.

vision navigation, 3D reconstruction, terrain matching, pose estimation

Citation: Zhu Z S, Su A, Liu H B, et al. Vision navigation for aircraft based on 3D reconstruction from real-time image sequences. Sci China Tech Sci, 2015, 58: 1196–1208, doi: 10.1007/s11431-015-5828-x

1 Introduction

The estimation of the position and attitude angle of a long-endurance aircraft is crucial for flight and path control. Fully autonomous navigation systems can improve an aircraft's anti-interference ability, reliability, and availability. Therefore, the estimation of the absolute position and attitude angle of an aircraft when there is a large drift in its inertial navigation system (INS) and when its global positioning system (GPS) is experiencing interference or is unavailable is an important research problem.

Vision navigation methods use passive image capture devices and computer vision methods to obtain the relative or absolute position and attitude angle information for aircraft platforms. These methods have the advantages of high autonomy, no accumulation errors, and comprehensive measurement parameters. These methods can be used independently or be combined with other navigation methods.

In terms of image texture information, the topography of an area changes little under normal conditions. Therefore, for relatively large terrain-relief areas, terrain-matching-based vision-navigation methods are more stable than scene-matching methods. The main advantage of image sequences obtained in a real-time flight is the ease of matching and tracking feature points with the overlapped region with time continuity. For a flight platform, one may choose a better intersection condition to achieve a better-reconstructed three-dimensional (3D) terrain. By introducing a reference terrain map, we can match the reconstructed 3D terrain over the flight region with the reference terrain map. This matching yields the correspondence between image points and 3D spatial points. Finally, the absolute position and orientation of the aircraft can be resolved by using pose estimation methods. Because most of the Earth's terrain

*Corresponding author (email: zzs2623@qq.com) information is readily available, this method can be implemented for fully autonomous navigation. These methods use passive imaging equipment instead of radar, laser ranging, and other active terrain measurement equipment. Further, they have considerable advantages in terms of energy consumption, payload weight reduction, and enhanced survival ability of the aircraft.

In the absence of any prior knowledge of flight parameters, the 3D reconstruction technique can only recover the 3D structure of the target scene. It cannot obtain the absolute scale scene structure unless the recovered 3D terrain map can be scaled to the reference terrain map. If the initial position cannot be determined, the navigation problem becomes a searching and matching problem between the reconstructed 3D terrain and the reference terrain under scale and perspective transformation conditions.

In this study, we deal with the problem of estimating the accurate position and attitude angle for a long-endurance flight. First, we employ the random sample consensus (RANSAC) and weighted estimator for a stable, accurate, and fundamental matrix estimation for 3D sequence reconstruction. Next, we propose a peak extractor and description method for terrain matching under a scale transformation condition. Finally, we improve the orthogonal iterative pose estimation method by using a weighted strategy.

2 Related work

In general, vision navigation methods can be divided into three types: scene matching-based approaches, integrated vision navigation methods with other sensor information, and autonomous navigation methods based on vision constraints between image feature points and ground feature points.

(1) Scene matching-based approaches. These methods calculate the position and attitude information by directly matching the real-time image and the pre-stored reference image. These methods are suitable for flat terrain and rich surface texture features of an area. For a low-altitude aircraft, by undulating the terrain area, we decreased the stability and accuracy of localization, particularly the accuracy of the attitude parameter. Moreover, when we consider the weather, climate change, and textures in different seasons, we find that a practical scene matching navigation method suffers from a further decline. Furthermore, an image changes under different situations such as different times, seasons, view angles, and heights, particularly in the case of mountains and hilly areas. Hence, it is difficult to directly match a reference image and a real-time image.

(2) Integrated vision navigation methods with other sensor information. By fusing the information obtained from an INS, radar, or other sensors, vision-based methods achieve pose estimation by using the aircraft speed, landmarks, and terrain elevation data. Methods such aircraft motion constraints are rather stringent, and some belong to the active measurement equipment, will reduce the capacity of the aircraft mobility and concealment.

(3) Autonomous navigation methods based on vision constraints between image feature points with ground feature points. This method uses 3D terrain information to eliminate the uncertainty of 3D reconstruction so as to obtain the aircraft position and orientation parameters by using the constraints of the terrain. Wang et al. [1] proposed a fusion of the GPS/INS navigation- and vision system-based approaches by using a laser altimeter and an optical flow analysis to achieve 3D scene reconstruction and a navigation information solver. Stevens et al. [2] used a real-time image sequence captured by an aircraft and described the 3D reconstruction ground terrain in absolute scale with the help of aircraft speed information. Then, they used the correspondence points between the reconstructed terrain and the reference terrain maps to determine the absolute aircraft position and posture. Sim et al. [3,4] proposed a combined navigation parameter estimation method by matching the reconstructed 3D terrain and the reference terrain maps to estimate the relative position and orientation of the aircraft and by using a combination of scene and terrain matching methods to obtain the absolute position and orientation information of the aircraft. However, these methods rely on an INS or a GPS to obtain information on different time motions and locations of the aircraft in order to obtain a real absolute scale of the terrain.

For terrain matching, Golden proposed the terrain contour matching (TERCOM) method [5]. The TERCOM method is based on the mean absolute difference (MAD) between the measured elevations and the map elevations of a given map profile. The missile position is calculated using a multi-state Kalman filter technique. After the determination of the TERCOM fix, the update is obtained by fusing with the Kalman filter equations. The TERCOM process requires a terrain map that is sufficiently unique in order to enable a missile to traverse from the launch to the target. Moreover, it requires the missile to fly in parallel to the columns of the TERCOM map. Behzad and Behrooz [6] introduced an iterative closest contour point (ICCP) from image registration to underwater terrain matching. Rodriguez and Aggarwal [7] proposed an approach that uses a reconstructed 3D terrain map and estimated the aircraft position by matching this reconstructed terrain map with a pre-stored DEM. A cliff map is used as a compact representation of the 3D surfaces.

When the aircraft velocity and orientation information is unknown, scale, translation, and rotation are observed between the reconstructed 3D terrain and the reference terrain map when the 3D terrain generation undergoes a different perspective transformation. Therefore, terrain contour matching methods such as MAD, mean square difference (MSD), cross-correlation, and contour-based methods are no longer applicable to the verification of existence of the scale transformation. Li et al. [8] presented a passive navigation method of terrain contour matching by reconstructing a 3D terrain from an image sequence. Control point tracking, key frame selection, and multiple view geometry were employed to accomplish 3D reconstruction. Then, terrain matching with a reference map provided the navigation information. In the follow-up study [9], they proposed a scale-invariant terrain matching method based on the oriented terrain surface features, by using the tangent of the terrain surface to construct the invariable features of the terrain and then matching the oriented terrain surface by calculating and comparing the distance of the feature vector. In [10], they improved the terrain matching method by using an invariant feature description of the 3D terrain for the terrain contour matching, which is robust for the similarity transformation of the terrain. The shape of the terrain is represented by an invariant feature vector based on the relative position distribution of the vertices.

With respect to pose estimation with the reference terrain map, Ronen [11,12] presented a method using DTM as a global reference to eliminate uncertainty in pose and motion estimation and self-localization. In the follow-up study [13], he established the direct constraints on the basis of illumination continuity and optimized the camera motion and scene structure. This method requires prior information about the pose of the camera in the first frame and assumes that the DTM can be linearized. It can guarantee that the algorithm does not fall prey to local extremism. Under the abovementioned conditions, the iteration method can obtain highly accurate position and attitude angle parameters.

3 Proposed method

The method proposed in this paper is mainly used for finding a solution for autonomous aircraft navigation when there are large drift errors in the INS or GPS jamming conditions and for estimating the initial position and pose without the use of INS and GPS. It focuses mainly on two aspects: the terrain matching problem under the scale transformation and view angle change condition, and the pose estimation problem when there are differences in the errors in correspondence between image points and 3D space points caused by terrain matching. By improving the stability and accuracy of terrain matching and pose estimation, the proposed navigation system becomes a completely independent navigation system, which can also be expanded to the scope of the other vision navigation systems.

The basic idea of 3D reconstruction terrain matching-based vision navigation is to use terrain features to establish the relationship between image points and the 3D reference points. The main procedure is shown in Figure 1. It can be divided into three main parts: 3D reconstruction from an image sequence, terrain matching, and navigation parameter solving.

3.1 Two-view reconstruction

Image sequence capture from a flying platform is used for easily satisfying an epipolar constraint relationship. We can use a two-view-based reconstruction technique to recover the terrain of the overlapped region from two images. In terms of the two-view geometry theory, the recovered terrain or targets can only maintain the shape and structure. Without any prior information of absolute scale, the estimated translation vector undergoes a scale transform with the actual translation vector, which implies that the recovered targets are not of the actual size. If the camera position or some control point information can be obtained, we can realize the absolute 3D reconstruction for the target position, shape, and size.

On a flight platform such as UAV, the mobility is large and the pitch angle and roll angle change significantly. Therefore, 3D reconstruction needs to ensure that the reconstruction frame selection has a large overlap and intersection angle. On the other hand, it is a more stable and accurate motion recovery method, namely the fundamental matrix estimation the problem.

The relationship between the fundamental matrix F and the correspondence points $\boldsymbol{x} \rightarrow \boldsymbol{x}'$ in images can be represented as follows: $\boldsymbol{x}'^{\mathrm{T}} \boldsymbol{F} \boldsymbol{x} = 0$. Given no less than 8 groups, the matching points can form linear equations.

Considering the mismatch and accuracy problems in feature point matching, in this study, we combine the RANSAC and adaptive weighted iteration methods for solving the fundamental matrix. This method first selects the inner point set by using RANSAC and applies this set to the epipolar line distance as the weighted function in the iteration procedure. This method can not only guarantee a robust solution but also achieve a highly accurate solution. On the basis of the nature of the fundamental matrix and the camera's internal parameters, the fundamental matrix can be decomposed into a relative projection matrix of the two cameras. Then, the 3D reconstruction can be realized by intersection.

The detailed steps of the algorithm are as follows.

(1) Using two-view feature extraction and matching, get the matching points $(\boldsymbol{x}, \boldsymbol{x}')$.

(2) Solving the fundamental matrix $\boldsymbol{F}$ with the matching points by the RANSAC method, get the subset of the inner points.

(3) Using the sum of the inner points to the epipolar line distances as the objective function and the individual points to the epipolar line as the weighting factor, and then, minimizing the objective function, get the optimized estimation of the fundamental matrix.

(4) Using the camera's intrinsic parameter K_1 and K_2, decompose the fundamental matrix into the relative rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$ of the two cameras.

(5) Using the relative camera motion $\boldsymbol{R}$ and $\boldsymbol{t}$, calculate

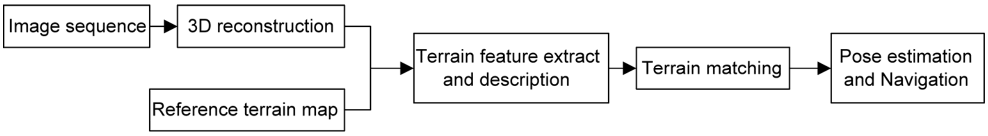
Image sequence
3D reconstruction
Reference terrain map
Terrain feature extract and description
Terrain matching
Pose estimation and Navigation

Figure 1 Main procedure for vision navigation system.

the projection matrix as follows:

$$\boldsymbol{P}_1 = K_1\left(\boldsymbol{I} \quad 0\right),\ \boldsymbol{P}_2 = K_2\left(\boldsymbol{R} \quad \boldsymbol{t}\right). \tag{1}$$

(6) Using the following equations, calculate the intersection of all the matching points:

$$\lambda \boldsymbol{x} = \boldsymbol{P}_1 \boldsymbol{x},\ \lambda' \boldsymbol{x}' = \boldsymbol{P}_2 \boldsymbol{x}. \tag{2}$$

3.2 Terrain matching based on shape description of mountain peak

3D terrain matching is the process of comparing a similarity and search process between the real-time reconstructed 3D terrain and the reference terrain maps. In order to design a property similarity measurement, we need to choose a compact and complete description. A terrain feature descriptor is the quantitative data for the local structure feature description, which should be able to fully reflect the shape and texture features of the local topographic area.

This method first detects the extreme local peaks in the terrain map, uses the elevation distribution to fit the surface of the peaks, and then uses the characteristic quadric to describe the shape. The characteristic quadric gives the initial similar transformation. We then adopt the region-based-matching method to find the correspondence matching points between the reconstructed 3D terrain and the reference terrain. This description has been proven to be invariant to similar transformations and is therefore applied to realize terrain matching between the reconstructed 3D terrain and the reference terrain.

3.2.1 Peak extraction

Unlike the MSER [14] method, the objective of the proposed method is terrain elevation. MSER extracts the feature region from the local texture distribution of the image; the texture of the same region has considerable uncertainty. However, the terrain elevation in the local range may not change without rules.

The MESR operator detects the maximally stable extreme region by the change in the water level. Inspired by the MSER method, we used the water level changes and the region growth method to detect the peak regions of the terrain. It is worth noting that the relative height of different peaks is constant and invariant to a similar transformation. During the detection of peak regions, the area of the previous peak region is specified by the subsequent lower peak region.

As shown in Figure 2(a), the terrain was inundated with water. When the water receded, the mountain was revealed gradually. Once the water level falls, one or more of the following four situations may occur: (1) No new peaks appear. (2) One or more new peaks appear. (3) The appeared peaks become larger. (4) Some peak areas merge.

In the detection of peak regions, the region growth method is used for connecting and marking the peak regions. Each independent peak is represented by a region centroid and area. Moreover, the terrain limited threshold will inevitably cause repetitive detection. Therefore, a distance threshold was employed to eliminate the repetitive peaks. Figures 2(b) and (c) show the terrain in 3D and grayscale. The peak regions are shown in Figure 2(d).

The use of the water level can lead to a correct detection of a peak only when the terrain is under the World Coordinate System (WCS). However, the reconstructed 3D terrain will suffer rotation transformation in the case of a relatively large viewpoint. Therefore, before the peak detection, we conduct a principle component analysis (PCA) to approximately match the reconstructed 3D terrain to that under the WCS.

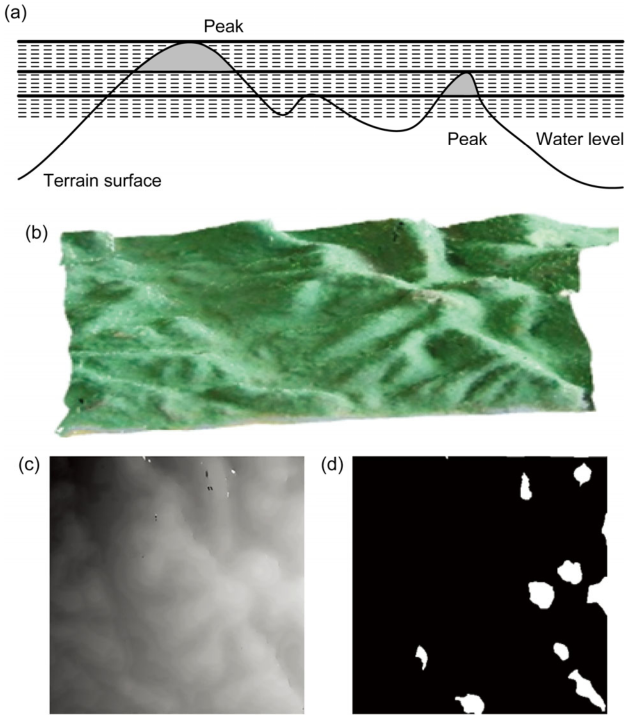

Figure 2 (Color online) The demo of peak extraction. (a) Water level change demo; (b) 3D terrain map; (c) terrain in grayscale map; and (d) peak region map.

The detailed steps of the peak extraction are as follows.

(1) Find the highest point and the lowest point of the terrain, and convert the topography into a grayscale image changing from 0 to 255.

(2) Use different thresholds for binary processing with the grayscale image. When the surface is greater than the threshold, the gray value will be marked as 255.

(3) Using the region growth method to connect the region marked as 255, calculate the centroid and area, and then, count the number of independent peaks.

(4) According to the requirement of peak counts, set the threshold count for Steps (2) and (3), and then, repeat them.

(5) Set the false alarm and confirm the threshold for the area of the peak region. When these areas reach this threshold, the centroid and area of the peaks will be added to the storage list.

(6) Employ the centroid distance of the peaks to determine the repeated peaks in the entire stored list. Set the minimum distance threshold. When the distance is less than this threshold, leave only the peaks located at the front of the list.

After extracting the peaks of the terrain by using this method, we can obtain the position and area information of the peaks, denoted as follows: $\boldsymbol{P}_i = \{x_i \quad y_i \quad A_i\}$, $i = 1, \cdots, n$.

Considering the practical application, we find that many peak areas are spread over the larger range of the terrain reference map and that the peak areas are relatively few in the reconstructed 3D terrain. The peak detection method detected the peak areas according to the sequential of the peak height. Therefore, we need to select a proper local range during the peak extraction for the terrain reference map.

3.2.2 *Peak description*

Among the peaks, the most common shapes are the ridge and the apex, as shown in Figures 3(a) and (b). Further, the main structure of the terrain can be simplified by using quadric surfaces. The approximate shape of these peaks can be represented by quadrics.

The quadrics can be described in a polynomial function form as follows:

$$\begin{aligned} \boldsymbol{F}(x,y,z) &= a_0x^2 + a_1y^2 + a_2z^2 + 2a_3xy + 2a_4xz + 2a_5yz \\ &+ 2a_6x + 2a_7y + 2a_8z + a_9 = 0. \end{aligned} \quad (3)$$

Then, it can be represented by a symmetric matrix as follows:

$$\begin{aligned} \boldsymbol{F}(x,y,z) &= [x \quad y \quad z \quad 1]\begin{bmatrix} a_0 & a_3 & a_4 & a_6 \\ a_3 & a_1 & a_5 & a_7 \\ a_4 & a_5 & a_2 & a_8 \\ a_6 & a_7 & a_8 & a_9 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \\ &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} = 0, \end{aligned} \quad (4)$$

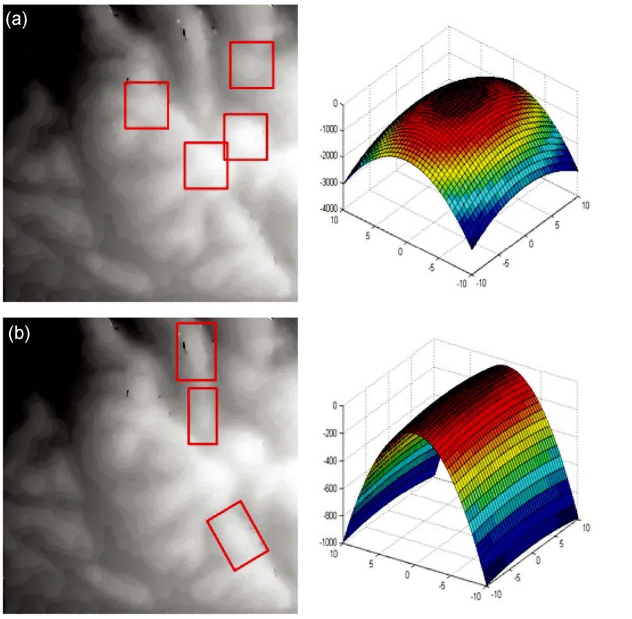

Figure 3 (Color online) Different types of peak shapes and quadric estimation: (a) apex area and (b) ridge area.

where $\boldsymbol{Q} = \begin{bmatrix} a_0 & a_3 & a_4 & a_6 \\ a_3 & a_1 & a_5 & a_7 \\ a_4 & a_5 & a_2 & a_8 \\ a_6 & a_7 & a_8 & a_9 \end{bmatrix}$ can identify the unique quadric, $\boldsymbol{Q}$ denotes a 4 × 4 symmetric matrix, and $\boldsymbol{x} = [x, y, z, 1]^{\mathrm{T}}$.

Since the matrix $\boldsymbol{Q}$ is symmetrical, it can be decomposed into the form $\boldsymbol{Q} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{U}$. Further, $\boldsymbol{U}$ represents an orthogonal matrix and $\boldsymbol{D}$ denotes a diagonal matrix. The elements of matrix $\boldsymbol{D}$ can be 0, 1, or −1 according to an appropriate scale transformation. Then, the quadrics can be expressed in several standard forms, such as ellipsoid, hyperboloid, cone, paraboloid, and cylinder surface.

Corresponding to the features of the 3D terrain, this quadratic component can reflect the contour and shape of the terrain. The ellipsoid and paraboloid surfaces are similar to gentle peak areas, and the hyperboloid and cone surfaces are similar to saddle peak areas.

$\boldsymbol{A} = \begin{bmatrix} a_0 & a_3 & a_4 \\ a_3 & a_1 & a_5 \\ a_4 & a_5 & a_2 \end{bmatrix}$ represents the quadratic component of $\boldsymbol{Q}$. The eigenvalues λ_x, λ_y, and λ_z represent the curvature in the different principal axes of the quadric surface, and the eigenvalues v_x, v_y, and v_z represent the directions of the principal axes. The main structure of a typical peak area and the quadric fitting are shown in Figure 4.

Upon the application of a point transformation $\boldsymbol{x}' = \boldsymbol{H}\boldsymbol{x}$, Eq. (4) becomes

$$\boldsymbol{x}'^{\mathrm{T}}\boldsymbol{Q}'\boldsymbol{x}' = (\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{Q}'(\boldsymbol{H}\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{Q}'\boldsymbol{H})\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} = 0. \quad (5)$$

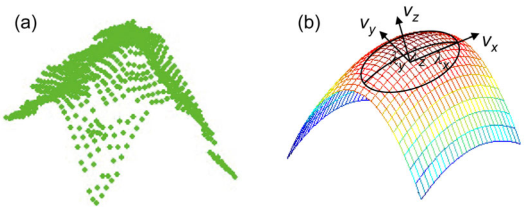

Figure 4 (Color online) Peak description: (a) peak area and (b) quadric fitting.

This results in the following transformation rule for a quadric: $\boldsymbol{Q}'=\boldsymbol{H}^{-\mathrm{T}}\boldsymbol{Q}\boldsymbol{H}^{-1}$. The transformation between $\boldsymbol{Q}$ and $\boldsymbol{Q}'$ can be expressed by saying that a quadric transforms a covariant. Further, point transformation $\boldsymbol{H}$ can be an arbitrary homogenous transformation; therefore, a similar transformation is included.

Assume that two peak areas $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ and the corresponding eigenvalues are $\boldsymbol{e}_1=(\lambda_{11}, \lambda_{12}, \lambda_{13})$ and $\boldsymbol{e}_2=(\lambda_{21}, \lambda_{22}, \lambda_{23})$; the corresponding eigenvectors are as follows:

$$\boldsymbol{V}_1=\begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{14} & v_{15} & v_{16} \\ v_{17} & v_{18} & v_{19} \end{bmatrix},\ \boldsymbol{V}_2=\begin{bmatrix} v_{21} & v_{22} & v_{23} \\ v_{24} & v_{25} & v_{26} \\ v_{27} & v_{28} & v_{29} \end{bmatrix}.$$

Since the SVD is an orthogonal decomposition, the eigenvectors give the direction along each axis of the quadric surface. The rotation matrix from $\boldsymbol{P}_1$ to $\boldsymbol{P}_2$ can be resolved by using $V_1V_2^{-1}$.

The 3D reconstruction method can only recover the 3D terrain with a scale factor up to the reference map. In the case of the scale transformation s, the eigenvalues of $\boldsymbol{Q}$ and $\boldsymbol{Q}'$ satisfied the following equation:

$$\frac{\lambda_{21}}{\lambda_{11}}=\frac{\lambda_{22}}{\lambda_{12}}=\frac{\lambda_{23}}{\lambda_{13}}=s^2. \tag{6}$$

Furthermore, the scale transformation has no influence on the quadric structure; the ratio between the curvatures of the different axes is constant, and the relationship can be expressed by the following equation:

$$\frac{\lambda_{11}}{\lambda_{13}}=\frac{\lambda_{21}}{\lambda_{23}},\frac{\lambda_{12}}{\lambda_{13}}=\frac{\lambda_{22}}{\lambda_{23}}. \tag{7}$$

In this model and under approximate conditions, quadric surface fitting was applied to the detected irregular surface areas, and the model parameters for the peaks were obtained. The characteristics of the peaks were represented by the quadric description. Furthermore, an iterative process can obtain a more stable peak position and description.

Select an appropriate standard coordinate system; the peak description can give the size information in the x and y directions as s_1 and s_2, and the angle information θ. These parameters can be expressed as follows:

$$s_1=\lambda_{11}/\lambda_{13},\ s_2=\lambda_{12}/\lambda_{13}. \tag{8}$$

In the previous PCA correction procedure, the angle of the pitch and roll is corrected according to the WCS, and the yaw angle represents the direction of the peaks, which can be expressed as follows:

$$\theta=\arctan(v_{11}/v_{12}). \tag{9}$$

By using the peak description method, we can obtain the position, scale, and direction information of the peaks; this can be denoted as follows:

$$\boldsymbol{P}_i=\{x_i \quad y_i \quad s_{1i} \quad s_{2i} \quad \theta_i\},\quad i=1,\cdots,n.$$

3.2.3 *Peak matching*

Since the peak extractor and description is based on the main structure of the peak, the peak description P_i is insufficient to identify the variety of peak shapes. Considering the search range, independence, and repeatability of feature description, we adopted a coarse-to-fine strategy to realize the matching procedure. First, the initial similar transformation between reconstructed 3D terrain is performed and the reference terrain map is estimated by using the peak descriptions. Then, an iterative matching method based on the local height distribution is employed to identify the correspondence matching points.

A similar transformation can be defined as follows:

$$\begin{bmatrix} x' \\ y' \end{bmatrix}=\begin{bmatrix} t_x & s_1r_{11} & s_2r_{12} \\ t_y & s_1r_{21} & s_2r_{22} \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}=\begin{bmatrix} a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{10}$$

where $\begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}=\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, which is given by the angle information, and $\begin{bmatrix} t_x \\ t_y \end{bmatrix}$ denotes the translation vector. Further, (x, y) and (x', y') are the coordinates of the image pairs.

We use a_0, a_1, a_2, b_0, b_1, and b_2 as the transformation parameters. Considering the noise and similar transformation, we can approximately describe the intensity variation between image pairs as follows:

$$\begin{aligned} g_1(x,y)+n_1(x,y)&=g_2(a_0x+a_1y+a_2,b_0x+b_1y+b_2) \\ &\quad+n_2(x,y), \end{aligned} \tag{11}$$

where n_1 and n_2 denote the random noise, and g_1 and g_2 represent the intensity values of the image pairs.

The least squares matching (LSM) method estimates movement by iteratively modifying the affine transformation and radiometric shift parameters to minimize the intensity difference in the correspondence area. The error equation for each pixel can be described as follows:

$$\Delta g = g_2(a_0x+a_1y+a_2, b_0x+b_1y+b_2)-g_1. \tag{12}$$

After linearization, the above equation can be rewritten as follows:

$$v = c_0da_0 + c_1da_1 + c_2da_2 + c_3db_0 + c_4db_1 + c_5db_2 - \Delta g, \tag{13}$$

where da_0, $\cdots$, db_2 denote the parameter deviation and the initial values given by the peak description, $\Delta g = g_2(x, y) - g_1(x, y)$, represent the intensity difference.

The parameters of the error equation can be expressed as follows:

$$\begin{aligned}&[c_0, c_1, c_2, c_3, c_4, c_5]\\ &=\left[x\frac{\partial g_2}{\partial x}, y\frac{\partial g_2}{\partial x}, \frac{\partial g_2}{\partial x}, x\frac{\partial g_2}{\partial y}, y\frac{\partial g_2}{\partial y}, \frac{\partial g_2}{\partial y}\right].\end{aligned} \tag{14}$$

According to the theory of LSM, for all the pixels within the match window, the objective function $s(x)$ of this optimized problem can be described as follows:

$$\min \boldsymbol{s}(x) = \sum_{i=1}^{n} \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{v}_i. \tag{15}$$

The linear Eq. (13) can be rewritten into a matrix form as follows:

$$\boldsymbol{v} = \boldsymbol{Ax} - \boldsymbol{b}, \tag{16}$$

where

$$\boldsymbol{A} = \begin{bmatrix} x\frac{\partial g_{21}}{\partial x_1} & y\frac{\partial g_{21}}{\partial x_1} & \frac{\partial g_{21}}{\partial x_1} & x\frac{\partial g_{21}}{\partial y_1} & y\frac{\partial g_{21}}{\partial y_1} & \frac{\partial g_{21}}{\partial y_1} \\ x\frac{\partial g_{22}}{\partial x_2} & y\frac{\partial g_{22}}{\partial x_2} & \frac{\partial g_{22}}{\partial x_2} & x\frac{\partial g_{22}}{\partial y_2} & y\frac{\partial g_{22}}{\partial y_2} & \frac{\partial g_{22}}{\partial y_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x\frac{\partial g_{2n}}{\partial x_n} & y\frac{\partial g_{2n}}{\partial x_n} & \frac{\partial g_{2n}}{\partial x_n} & x\frac{\partial g_{2n}}{\partial y_n} & y\frac{\partial g_{2n}}{\partial y_n} & \frac{\partial g_{2n}}{\partial y_n} \end{bmatrix},$$

$$\boldsymbol{x} = [\mathrm{d}a_0 \quad \mathrm{d}a_1 \quad \mathrm{d}a_2 \quad \mathrm{d}b_0 \quad \mathrm{d}b_1 \quad \mathrm{d}b_2]^{\mathrm{T}},$$

$$\boldsymbol{b} = [g_{21}-g_{11} \quad g_{21}-g_{12} \quad \cdots \quad g_{2n}-g_{1n}]^{\mathrm{T}}.$$

When $n > 6$, Eq. (16) is an over-determined equation and the solution of $\boldsymbol{x}$ can be computed by using the conventional linear least squares method:

$$\boldsymbol{x} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \tag{17}$$

This iterative matching method iteratively estimates the displacement by minimizing an error function on the basis of the space distance within the matching window.

The detailed implementation steps are as follows:

(1) Detect the peak areas in a real-time recovered terrain, and give the shape description to the peaks.

(2) Detect the peak areas in the reference terrain map, and give the shape description to each peak; this procedure can be performed in advance, and the peak extractor should be taken block by block.

(3) Use the peak description information (scale and direction of the peak) to normalize the match area, and find the potential matching for each peak in the real-time terrain.

(4) Adopt the iterative matching method to confirm the match results.

3.3 Pose estimation

Since the reference terrain map provides the 3D coordinates of the ground points, with the corresponding image points, the estimation of the camera position and attitude angle becomes a standard pose estimation problem. When there are more than 6 non-planar or 4 coplanar feature points, the external camera parameters can be resolved by using a linear solution. Considering that the imaging process is nonlinear, we find that the accuracy of the direct linear solution is not high. In this study, the initial external parameters were obtained by using a linear solution first, and then, a nonlinear iteration was employed to optimize the initial solution.

The basic principle of vision-based navigation is to use the correspondence between image points and 3D space points to estimate the position and attitude angle of the onboard camera. The schematic representation is shown in Figure 5, where O denotes the optical center of the camera on the aircraft, I_1–I_n represent the image points, and T_1–T_n indicate the 3D position of the object point on the ground. Using the image coordinates and the world coordinates of the corresponding points, the central perspective projection model can solve the position and attitude angle of the aircraft.

3.3.1 *Initial pose estimation*

The central perspective projection model can be represented as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & C_x & 0 \\ 0 & f_y & C_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ 0^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} \mathrm{X} \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{18}$$

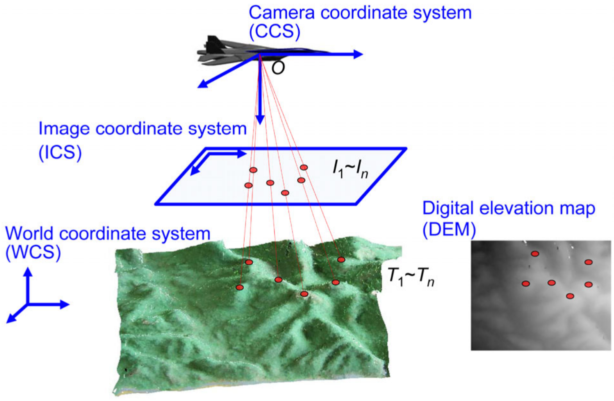
Camera coordinate system (CCS)
O
Image coordinate system (ICS)
I_1~I_n
Digital elevation map (DEM)
World coordinate system (WCS)
T_1~T_n

Figure 5 (Color online) The principle of vision-based navigation.

where u and v denote the image coordinates of the object points; and X, Y, and Z represent the 3D location in world coordinates. Further, f_x, f_y, C_x, and C_y denote the camera focal length and principal point coordinates; the internal camera parameters can be calibrated in advance. $\boldsymbol{R}$ and $\boldsymbol{t}$ represent the rotation matrix and the translation vector from the WCS to the CCS, respectively, which have 6 degrees of freedom. Each corresponding point can provide two equations; therefore, the problem becomes solvable if there are more than 6 pairs of corresponding points. Thus, the collinearity equation can be written as follows:

$$\begin{cases} Xr_0+Yr_1+Zr_2+t_X-\dfrac{u-C_x}{f_x}(Xr_6+Yr_7+Zr_8+t_Z)=0,\\ Xr_3+Yr_4+Zr_5+t_Y-\dfrac{v-C_y}{f_y}(Xr_6+Yr_7+Zr_8+t_Z)=0,\end{cases}\tag{19}$$

where $\boldsymbol{R}=\begin{bmatrix} r_0 & r_1 & r_2\\ r_3 & r_4 & r_5\\ r_6 & r_7 & r_8\end{bmatrix}$, and $\boldsymbol{t}=\begin{bmatrix} t_X\\ t_Y\\ t_Z\end{bmatrix}$.

Normalize the parameters by t_Z, and let $s_i = r_i/T_z$ (i = 0, $\cdots$, 8), $s_9 = T_x/T_z$, and $s_{10} = T_y/T_z$. Therefore, the above equation can be rewritten as the following equation:

$$\begin{cases} Xs_0+Ys_1+Zs_2+s_9-\dfrac{u-C_x}{f_x}(Xs_6+Ys_7+Zs_8)\\ =\dfrac{u-C_x}{f_x},\\ Xs_3+Ys_4+Zs_5+s_{10}-\dfrac{v-C_y}{f_y}(Xs_6+Ys_7+Zs_8)\\ =\dfrac{v-C_y}{f_y}.\end{cases}\tag{20}$$

By solving these linear equations, we can obtain s_0–s_{10}. According to the constraint $r_6^2+r_7^2+r_8^2=1$, $t_Z=\sqrt{1/(s_6^2+s_7^2+s_8^2)}$. The rotation matrix R and translation vector $\boldsymbol{t}$ can be resolved naturally. According to the definition of the attitude angle, the yaw, pitch, and roll angles can be decomposed by using the rotation matrix.

3.3.2 Adaptive weighted orthogonal iteration pose estimation

The fast and global convergence orthogonal iterative (OI) algorithm is one of the optimal methods for real-time pose estimation. Lu et al. [15] proposed the OI algorithm, which is a pose estimation algorithm for a single camera, in 2000.

With the knowledge of image points and their correspondence space points, the OI algorithm minimizes the space co-linearity error and calculates the rotation matrix and the translation vector simultaneously. However, the optimization of this method is based on the least squares theory; only when the errors follow the Gaussian distribution can we achieve the optimal solution. The corresponding image points and 3D terrain points obtained by terrain matching will inevitably suffer from errors. Moreover, the error may not follow the Gaussian distribution because of the terrain relief and undulate. In order to achieve a high-precision, robust pose estimation for the aircraft, we developed a weighted function for each match point and adjusted the weighted factor during the iteration.

Given a set of non-collinear 3D coordinates of object points $\boldsymbol{p}_i = (x_i, y_i, z_i)^{\mathrm{T}}$, i = 1, $\cdots$, n and $n \geqslant 3$ in the WCS, we can describe the corresponding $\boldsymbol{q}_i=\left(x_i', y_i', z_i'\right)^{\mathrm{T}}$ in the CCS as follows:

$$\boldsymbol{q}_i=\boldsymbol{R}\boldsymbol{p}_i+\boldsymbol{t},\tag{21}$$

where $\boldsymbol{R}=\left(\boldsymbol{r}_1^{\mathrm{T}}, \boldsymbol{r}_2^{\mathrm{T}}, \boldsymbol{r}_2^{\mathrm{T}}\right)^{\mathrm{T}}$, $\boldsymbol{t}=(t_x, t_y, t_z)$ denote the rotation matrix and the translation vector. Let the image point $\boldsymbol{v}_i = (u_i, v_i, 1)^{\mathrm{T}}$ be the projection of $\boldsymbol{p}_i$ on the normalized image plane. The collinearity equation can be expressed as follows:

$$\boldsymbol{R}\boldsymbol{p}_i+\boldsymbol{t}=\boldsymbol{V}_i\left(\boldsymbol{R}\boldsymbol{p}_i+\boldsymbol{t}\right).\tag{22}$$

Therefore, the object space collinearity error can be derived as follows:

$$\boldsymbol{e}_i=\left(\boldsymbol{I}-\boldsymbol{V}_i\right)\left(\boldsymbol{R}\boldsymbol{p}_i+\boldsymbol{t}\right),\tag{23}$$

where $\boldsymbol{V}_i=\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}/\left(\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_i\right)$ represents the observed line-of-sight projection matrix. This error can be interpreted as the distance from the point to the line of sight.

The objective function of the OI algorithm is to minimize the object space collinearity errors. We added a weighted function to the objective function. The objective function can be expressed as follows:

$$E(\boldsymbol{R},\boldsymbol{t})=\sum_{i=1}^{n}\left\|w_i e_i\right\|^2=\sum_{i=1}^{n}\left\|w_i(\boldsymbol{I}-\boldsymbol{V}_i)(\boldsymbol{R}\boldsymbol{p}_i+\boldsymbol{t})\right\|^2,\tag{24}$$

where w_i denotes the weighted factor. Since the goal is to minimize the sum of the point errors, we use the Huber weighted function with the distance from the point to the line of sight.

The weighted function is defined as follows:

$$w_i=\begin{cases} 1, & |e_i|\leqslant\sigma,\\ \sigma/|e_i|, & \sigma<|e_i|\leqslant 3\sigma,\\ 0, & 3\sigma<|e_i|,\end{cases}\tag{25}$$

where σ=1.4826(1+5/(n–7)) with the median $|e_i|$ in common and n denotes the number of corresponding points.

In order to reduce the impact of the larger re-projection error matching points, the Huber weighted function was employed to adjust the weight factor. The small re-projection error matching points were regarded as interior points by the Huber weighting function and assigned a weight of 1. In a practical application, the weighting function should be further designed on the basis of the engineering requirements.

4 Simulations and experiments

In order to evaluate and verify the performance of the proposed approach for vision-based navigation, a series of simulations and flight experiments were carried out.

4.1 Flight imaging simulation

The discrete 3D terrain data were stored by DEM in the standard manner. The undulating and disorder terrain surface data could not be represented by an arbitrary formula. In order to generate an image under a perspective projection, we proposed a flight imaging simulation method based on line-to-plane intersection with iterative searching. A set of DEM data with texture was employed to simulation the 3D scene, and the aircraft position and attitude and the camera's intrinsic and external parameters were given as required. Assume that the optical center of imaging $\boldsymbol{P}_c = (x_c, y_c, z_c)$ is coincident with the aircraft under the DEM coordinate system and the camera external parameter is $\boldsymbol{R}_c$. Then, the relationship between the image coordinates and the DEM coordinate system can be derived as follows:

$$\boldsymbol{X}_c = \boldsymbol{R}_c(\boldsymbol{X}_w - \boldsymbol{P}_c). \tag{26}$$

The line direction vector given by image point (u, v) and the camera's principal point can be expressed as follows:

$$\begin{bmatrix} l \\ m \\ n \end{bmatrix} = \begin{bmatrix} u-u_0 \\ v-v_0 \\ f \end{bmatrix}, \tag{27}$$

where (u_0, v_0) denotes the principal point and f represents the focal length. The ray from the optical center to the image point and intersection the 3D terrain surface at $\boldsymbol{P}' = (x', y', z')$ can be expressed as follows:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} + k \begin{bmatrix} l \\ m \\ n \end{bmatrix}, \tag{28}$$

where k denotes the depth from the optical center to the terrain surface. Because the 3D terrain data are composed of 3D discrete points and cannot be described by a parametric equation, the ray intersection with the 3D terrain data cannot be solved directly with an analytical method and can only use the search method.

With the knowledge of the range of elevation distribution, the initial corresponding position can be calculated by presetting a reference plane. Then, an iterative searching method is employed to minimize the point-to-line distance and to find the best intersection position. The reference plane intersection and iterative procedure is shown in Figure 6.

We simulate the flight and imaging procedure by using a group of 3D landscape maps as the test dataset, including 3D terrain maps and texture data. The range of elevation is 1900–4600 m, and the grid size is 30 m × 30 m. The simulated image size is 640 × 480 pixels, and the focal length is 400. The simulation environment is shown in Figure 7(a). The DEM in grayscale is shown in Figure 7(b). Two simulation images from different view angles and locations are shown in Figures 7(c) and (d), respectively.

4.2 Reconstruction precision analysis

We select two areas with different undulations among the terrain and compare the reconstruction terrain data with the real reference terrain data. Because the image matching method is based on the assumption of an affine transform, the matching precision of the planar area is higher than that of an area with larger undulation. The results are shown in the following figures.

4.2.1 Simulation parameters

The simulated image size is 640 × 480 pixels, the focal length is 400, and the principle point is at (9000 m, 8500 m, 8000 m). The overall range of the terrain height is from 1900 m to 4600 m, and the grid size is 30 m × 30 m. The planar area height ranges from 3440 m to 3510 m, and the

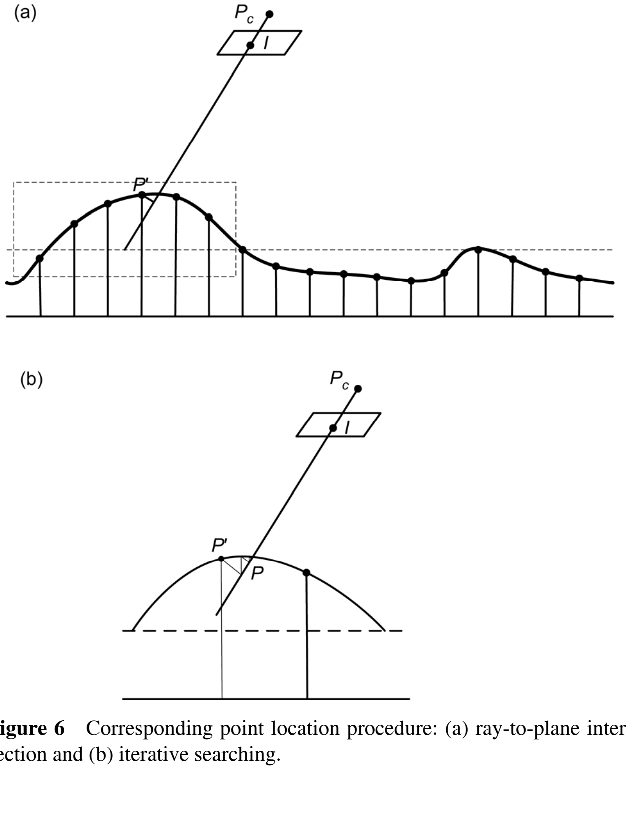

Figure 6 Corresponding point location procedure: (a) ray-to-plane intersection and (b) iterative searching.

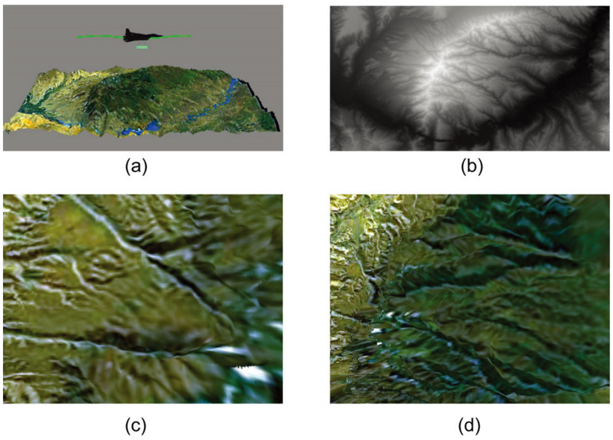

Figure 7 (Color online) Flight imaging simulation on DEM: (a) simulation environment; (b) DEM in grayscale; (c) view angle 1; and (d) view angle 2.

peak area height ranges from 3863 m to 4150 m. The different areas are shown in Figure 8.

4.2.2 Reconstruction precision evaluation

By introducing the baseline information, we can obtain the absolute space points by 3D reconstruction. Then, we give the errors between the reconstructed space points with the reference terrain map.

As we can see from Figure 9, the errors in the x and y directions are smaller than those in the z direction, and the errors in the z direction are related to the degree of variation on the terrain surface. Moreover, the errors are related to the texture feature above the surface; therefore, the errors in the x, y, and z directions seem to vibrate periodically.

4.3 Terrain matching experiments

In order to evaluate the performance of our approach for peak matching, we used a series of geometry transformations to verify them. Through the peak extractor, description, and iterative matching, we obtained the matching peaks. The matching results under scale, rotation, and similar transformations are shown in Figure 10.

As long as the reconstructed 3D terrain and the reference terrain map had at least one matched peak, we could estimate the similar transformation relationship between the reconstructed 3D terrain and the reference terrain map, and determine the location of the reconstructed 3D terrain in the reference terrain map. Furthermore, the region-based matching method was employed to improve the accuracy of terrain matching.

The reconstructed 3D terrain and the match results of the proposed method are shown in Figure 11. As can be seen from the figure, the recovered terrain retained the shape and structure of the real terrain, but the scale and distribution of the elevation changed significantly. The peaks, ridges, and other contours became relatively blurred. Further, the matching results demonstrate that the proposed matching method can overcome the noise, blur, and scale change factors, and obtain accurate, stable peak matching.

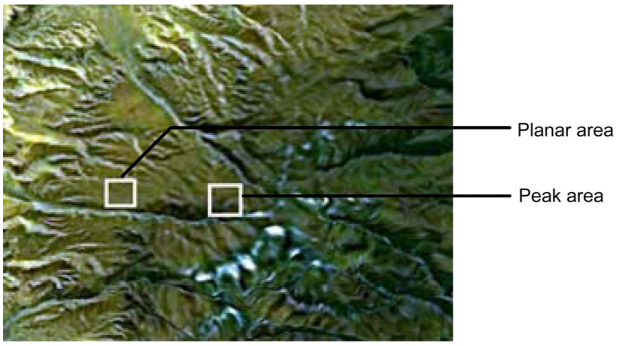

Figure 8 (Color online) Two selected areas for reconstruction precision evaluation.

4.4 Pose estimation comparisons

Because of the 3D reconstruction and terrain matching, the corresponding image points and space points had errors. The relationship between the two view angles was obtained by image feature point matching, and a dense disparity map was generated using stereo matching. Not only the image feature, region matching will introduce errors, but also the terrain matching. Because of the terrain relief and undulation, the matching errors were not always uniform. As shown in Figure 12(a), most of the errors were within 50 m, but larger errors could also appear. As can be seen from Figure 12(b), this adaptive weighted orthogonal iterative pose estimation algorithm could quickly perform iterative convergence and significantly reduce the object space collinearity errors.

4.5 Vision navigation experiments

Based on the previous DEM environments, we simulated the trajectory to test the adaptability and accuracy of the proposed algorithm. The flight trajectory and attitude angle were given by Eqs. (29) and (30). The unit of time was seconds and x, y, and z denote the position of the aircraft on the ground coordinate system using the unit of meters. A_x, A_y, and A_z represent the pitch, roll, and yaw angle, respectively, in the unit of degrees.

The position trajectory can be expressed as follows:

$$\begin{cases} x = 6000 + 30 \cdot t, \\ y = 4500 + 200 \cdot \sin \omega_1 t, \\ z = 8000 + 200 \cdot \sin \omega_2 t. \end{cases} \tag{29}$$

The attitude angle can be expressed as follows:

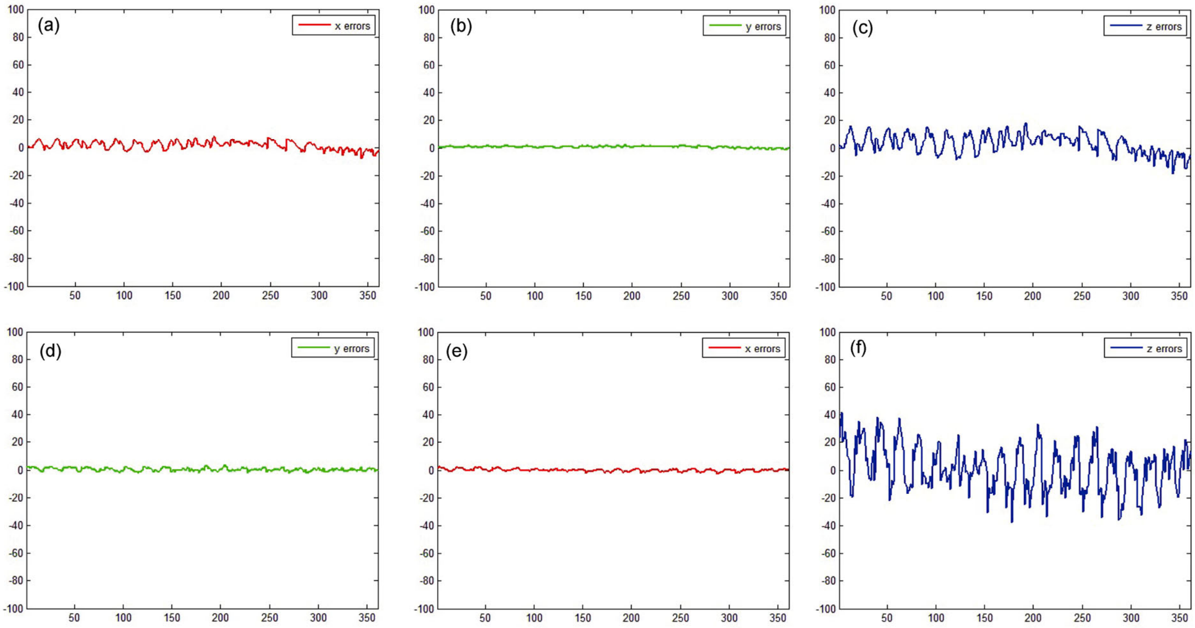

Figure 9 (Color online) The planar area: (a) *x*-direction errors; (b) *y*-direction errors; and (c) *z*-direction errors. The peak area: (d) *x*-direction errors; (e) *y*-direction errors; and (f) *z*-direction errors.

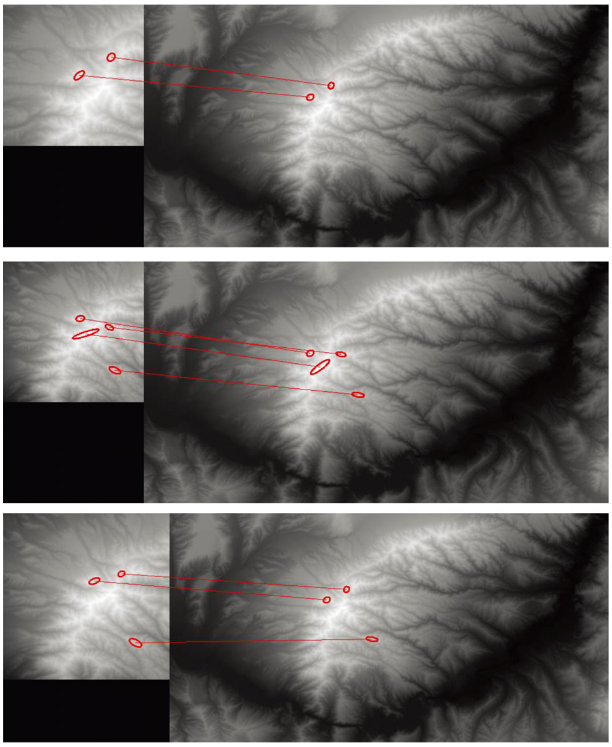

Figure 10 (Color online) Peak matching results: (a) scale transformation; (b) rotation transformation; and (c) similar transformation.

$$\begin{cases} A_x = 3\cdot\cos\omega_2 t, \\ A_y = 3\cdot\sin\omega_1 t, \\ A_z = 3\cdot\cos\omega_2 t, \end{cases} \tag{30}$$

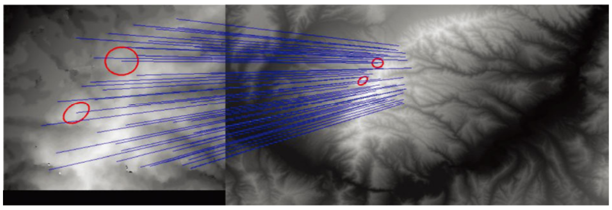

Figure 11 (Color online) Terrain matching results.

where $\omega_1 = \pi/300$ and $\omega_2 = \pi/300$.

In this study, we used the terrain matching method described in Section 3.2 and applied the adaptive weighted orthogonal iteration pose estimation method to obtain the position and attitude angle parameters, as shown in Figure 13. As can be seen from the graph, this method can obtain stable position and attitude parameters under flight maneuvering conditions and can guarantee an aircraft position error of less than 30 m and an attitude angle error of less than 0.5°. Further, under different terrain relief and undulation conditions, the pose estimation accuracy changes.

4.6 Vision/inertial integrated navigation experiments

In a vision/inertial integrated navigation system, the position measured by the vision-based method was applied to correct the inertial integration error by using a Kalman filter. The flight trajectory and attitude angle were consistent with the values obtained in the previous section.

The parameters for the INS simulation were as follows:

(1) initial error of the accelerometer = 0.5 mg (1σ)

(2) random error = 0.25 mg (1σ)

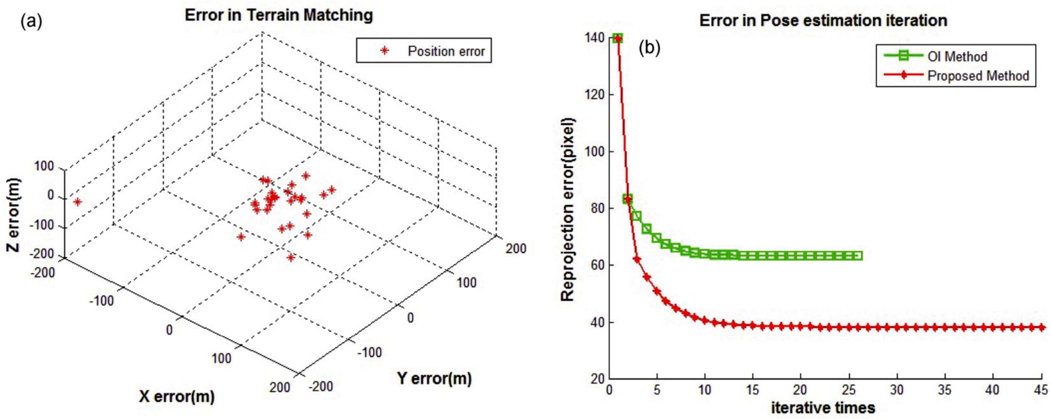

Figure 12 (Color online) Experiment for comparison: (a) terrain match errors and (b) iteration in pose estimation.

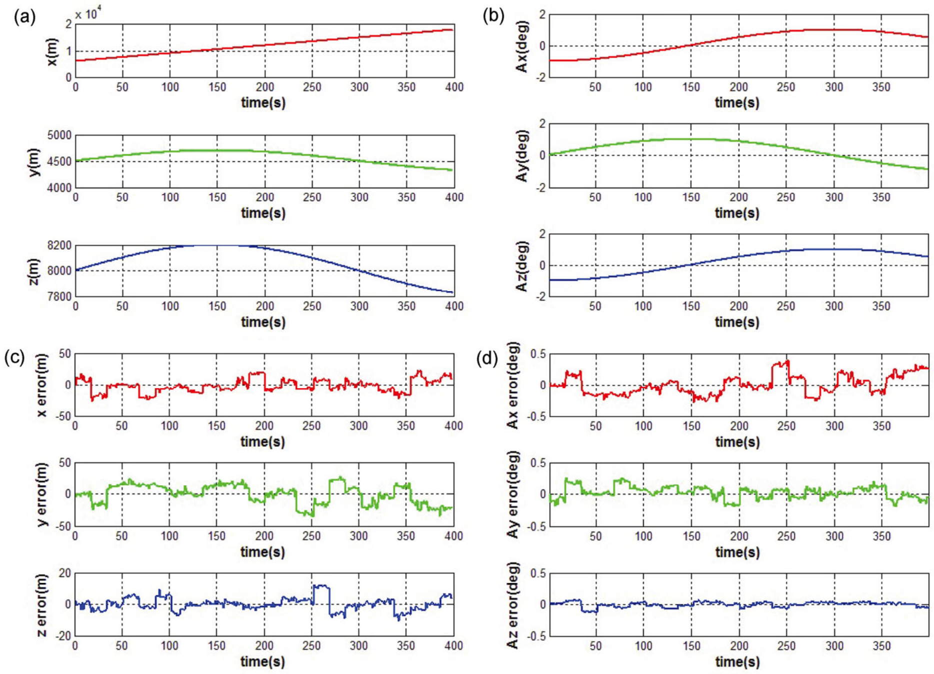

Figure 13 (Color online) Pose estimation by the proposed method: (a) position trajectory; (b) attitude angle; (c) position errors; and (d) attitude angle errors.

(3) random walk error = 0.25 mg (1σ)

(4) attitude angle error of the aircraft = 0.1°

(5) initial position error = 50 m

(6) initial velocity error = 1.2 m/s

(7) flight duration = 350 s

(8) inertial frequency = 100 Hz.

(9) vision measurement frequency = 1 Hz

Vision measurement from 0 s began to intervene and correct the inertial velocity with the Kalman filter. The position and velocity errors of the aircraft are shown in Figure 14. As can be seen from Figure 14, with the vision measurement data, the position and velocity errors converge to zero quickly. The position error of the aircraft can be maintained within 20 m, and the velocity error can be maintained within 0.25 m/s.

5 Conclusion

This paper gives deep insights on terrain matching and pose estimation for vision-based navigation problems. The peak extractor and description gives a robust matching result for terrain matching, and the adaptive weighted orthogonal iteration pose estimation method is robust to the noise caused by feature matching, stereo matching, and terrain matching errors. Simulations of the fly trajectory on the DEM and onboard imaging for estimating the position and attitude angle by the proposed method demonstrate promising performance in terms of providing robust and accurate navigation parameters. Further research will focus on the terrain matching of different perspectives and different terrain fea-

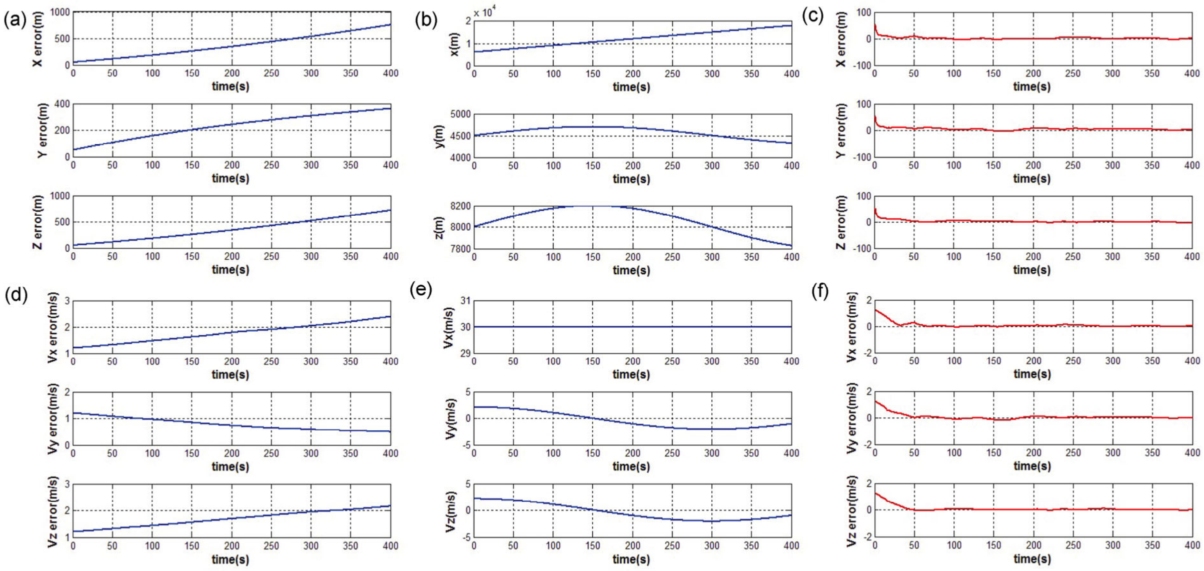

Figure 14 (Color online) Comparisons of aircraft position and velocity errors. (a) Flight path; (b) position error in INS; (c) position error in vision/INS; (d) flight velocity; (e) velocity error in INS; and (f) velocity error in vision/INS.

tures, and applying and improving the vision-based navigation system in practical applications.

This work was supported by the National Basic Research Program of China (“973” Project) (Grant No. 2013CB733100).

1 Wang J L, Garratt M, Lambert A, et a1. Intergration of GPS/NS/VISION sensors to navigate unmanned aerial vehicles. In: The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences, 2008. 963–969

2 Stevens M R, Snorrason M, Eaton R, et a1. Motion imagery navigation using terrain estimates. In: Proceedings of the 17th International Conference on Pattern Recognition, Cambridge. 2004, 1: 272–275

3 Sim D G, Park R H, Kim R C, et al. Integrated position estimation using aerial image sequences. IEEE T Pattern Anal, 2002, 24: 1–18

4 Sim D G, Park R H. Localization based on DEM matching using multiple aerial image pairs. IEEE T Image Process, 2002, 11: 52–56

5 Golden J P. Terrain contour matching (TERCOM): A cruise missile guidance aid. Image Process Missile Guid, 1980, 238: 10–18

6 Behzad K P, Behrooz K P. Vehicle localization on gravity maps. In: Proceedings of SPIE-The International Society for Optical Engineering, Orlando, Florida, 1999. 3693: 182–191

7 Rodriguez, J J, Aggarwal J K. Matching aerial images to 3-D terrain maps. IEEE T Pattern Anal, 1990, 12: 1138–1149

8 Li L C, Yu Q F, Shang Y et al. A new navigation approach of terrain contour matching based on 3-D terrain reconstruction from onboard image sequence. Sci China Tech Sci, 2010, 53: 1176–1183

9 Li L C, Yuan Y, Gui Y, et al, Scale Invariant terrain matching based on oriented terrain-surface feature. Comput Eng Appl, 2010, 46: 236–239

10 Li L C, Yuan Y, Li Y, et al, Invariant feature vector description for 3D terrain and its application to terrain contour matching. Acta Aeronaut Astronaut Sinca, 2009, 30: 2143–2148

11 Lerner R, Rivlin E, Rotstein H P. Pose and motion recovery from feature correspondences and a digital terrain map. IEEE T Pattern Anal, 2006, 28: 1404–1417

12 Lerner R, Rivlin E, Rotstein P H. Pose estimation using feature correspondences and DTM. In: Image Processing International Conference on Image Processing, Singapore, 2004. 2603–2606

13 Lerner R, Rivlin E. Direct method for video-based navigation using a digital terrain map. IEEE T Pattern Anal, 2011, 33: 406–411

14 Matas J, Chum O, Urban M, et al. Robust wide-baseline stereo from maximally stable extremal regions. Image Vision Comput, 2004, 22: 761–767

15 Lu C P, Hager G D, Mjolsness E. Fast and globally convergent pose estimation from video images. IEEE T Pattern Anal, 2000, 22: 610–622