A method of farmland 3D terrain reconstruction based on UAV sequence images

Cheng Xu1,2, Chenglong He2, Kun Liu3
1Research Institute of UAV, Nanjing University of Aeronautics and Astronautics, Nanjing, China
2State Key Laboratory of Satellite Navigation System and Equipment Technology, Shijiazhuang,050081, China
3College of Electronic Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Abstract. When the plant protection UAV on farmland spraying, the UAV need to always maintain a relative height of 1-2 m away from the crops to ensure uniform and efficient spraying of pesticides. At present, the commonly used high-fixing schemes, such as GPS, barometer, ultrasound and so on, there are some deficiencies. Aiming at this problem, a method of reconstructing 3D terrain of farmland by UAV electro-optical payload taking sequence images is proposed, which can achieve the goal of UAV height-fixing. Firstly, the feature points selection module is established. The SILC image segmentation algorithm is adopted to segment the image and the centroid points of all the superpixel blocks are calculated. Then, the corner points are extracted by using Harris corner detector and these two parts of points are mixed together as the feature points. Secondly, the feature point matching module based on SAD algorithm is proposed. Finally, a three-dimensional reconstruction algorithm according to monocular sequence images, UAV position, attitude and camera pointing angle is derived. The results of the experiments demonstrated that this method can quickly and effectively reconstruct farmland terrain, which can provide a data base for UAV high-fixing and has important application value.

1. Introduction
In recent years, unmanned aerial vehicles (UAVs) have made rapid development in many professional applications, in which agricultural plant protection UAV gains the most concern. For paddy fields, hills, mountains and other areas in which ground agricultural machinery and fixed-wing manned aircraft cannot work, plant protection UAV has unique advantages. One plant protection UAV is equivalent to 30 humans in operational efficiency, so it is welcomed by the farmers. However, there are some problems that plague UAV manufacturers all the time, such as flight control systems, pesticide spray system and flight platform. Among them, high-fixing for the plant protection UAV is the most disturbing. UAVs need to always maintain a relative height of 1 to 2 meters away from the crops to ensure uniform and efficient spraying of pesticides in the operation, which is known as the “terrain following” in industry terminology. But in the actual operation process, uneven farmland, weather, light and other factors also have great impact on plant protection UAVs.

At present, the traditional high-fixing schemes include GPS, barometer and ultrasonic wave [1-3]. These traditional height-definition schemes have many limitations, and the consequences include leakage and re-spray to influence the spray effect in the operation by plant protection UAV, which
even lead to explosion accident due to height drop. Therefore, it is very necessary to propose a reasonable and applicable plan for the height set of plant protection UAV.

Stereo vision technology can reconstruct the depth information of the imaging area, so as to provide theoretical basis and ways for height set by vision-based unmanned aerial vehicle. Stereo matching is to find the one-to-one correspondence between the pixels of projected images in the same spatial scene under different viewpoints, which is the key step to get the depth information of the scene [4].

Structure from motion (SfM) is a method of reconstructing three-dimensional structure by image matching, which is the key technique in 3D reconstruction. In the reconstruction of motion restoration, the most representative and influential work is Photo tourism system by Snavely [5], in which the published kernel system Bundler has completed the three-dimensional point coordinates construction from two-dimensional images of different viewpoints and the restoration of camera intrinsic parameters, position, and orientation.

If the Bundler algorithm is applied to reconstruct the UAV images directly, the efficiency of algorithm is low for the UAV images with large data volume. In actual use, the camera intrinsic parameters information is usually available, accompanied with some other priori information with low accuracy, such as flight route, UAV navigation data and stable platform posture, and making full use of these auxiliary information can improve the accuracy and efficiency of reconstruction.

Compared with the binocular camera, the monocular camera is more portable and easy to install on the unmanned aerial vehicle. Aiming at the above problems, a 3D terrain reconstruction algorithm based on monocular sequence images is proposed for the UAVs. This paper is organized as follows.

Section II gives an overview on the proposed scheme firstly, and then Section III introduces the feature point selection and matching algorithm and the three-dimensional reconstruction model in detail. Experiment results are given Section IV. Section V concludes the work.

2. System structure

We built a custom UAV (Figure 1) using a QAV280 quad-rotor UAV frame, a AMP autopilot equipped with GPS and compass, a 915 MHz telemetry radio to the ground station computer, a RC832 image receiver, a TS832 image transmitter and 1 cell LIPO batteries which sustained a 15-min flight.

Due to payload restrictions, the UAV is equipped with a single downward-looking camera (SONY 4140 CCD). After the image is taken by the camera, it is transmitted to the ground station through the image transmitter. The design aims to calculate the three-dimensional coordinates of the plant in real time by the ground station software according to the transmitted image sequences and the flight data of the UAV, which can provide data support for the flight by certain height.

From the two aspects of data accuracy and real-time calculation, quasi-dense matching is selected for 3D reconstruction. The overall block diagram is shown in Figure 2. Firstly, the SILC (simple linear iterative clustering) image segmentation algorithm is adopted to segment the initial image \(I_p\) and the centroid coordinates of all the superpixel blocks are calculated to obtain the point set \(P_1\). Then, the corner points of \(I_p\) are extracted by using Harris corner detector to obtain the point set \(P_2\). Then, \(P_1\) and \(P_2\) are mixed together to get feature point set \(P_p\). SAD (sum of absolute differences) algorithm is applied to detect the matching points of \(P_p\) in the current image \(I_c\) to obtain the point set \(P_c\). \(I_p\) and \(I_c\) do not coincide completely, so \(P_c < P_p\). Finally, the three-dimensional coordinates of homologous points are calculated according to three-dimensional reconstruction model.
3. Proposed scheme

3.1. Feature points selection module

Segmentation refers to the process of subdividing a digital image into a plurality of image subregions (a collection of pixels, also referred to as superpixel). The superpixel is a small area consisting of a series of neighboring pixels with similar characteristics such as color, brightness, and texture. Most of these small areas retain useful information for further image segmentation, and generally do not destroy the boundaries of objects in the image.

Simple linear iterative clustering (SLIC) is an adaptation of k-means for superpixel generation [6]:

In this paper, we use SILC algorithm to segment the initial image firstly. When we obtained a series of superpixels, the next step is calculate the centroid coordinate of every superpixel.

In image processing, computer vision and related fields, an image moment is a certain particular weighted average (moment) of the image pixels’ intensities, or a function of such moments, usually chosen to have some attractive property or interpretation. Image moments are useful to describe objects after segmentation. Simple properties of the image which are found via image moments include area (or total intensity), its centroid, and information about its orientation. In this case, we use image moments to find the centroid of each superpixel and obtain the point set \( P_1 \).

Obviously, through the calculation of first two steps, we can obtain some feature points, but these points can not describe the edge information. A corner can be defined as the intersection of two edges. A corner can also be defined as a point for which there are two dominant and different edge directions in a local neighbourhood of the point.

The Harris corner detector is a popular corner point detector due to its strong invariance to: rotation, scale, illumination variation and image noise. The Harris corner detector is based on the local auto-correlation function of a signal: where the local auto-correlation function measures the local changes of the signal with patches shifted by a small amount in different directions. Here, we use Harris corner detector to detect corner and obtain the point set \( P_2 \).

Then, \( P_1 \) and \( P_2 \) are mixed together to get feature point set \( P_p \).

3.2. Feature points matching module

The design choice has been made to employ a \( N_t \times N_t \) area taken around the feature point \((x, y)\) as a template. This module attempts to find a match for the template within the current image. To reduce
the search space for this match the search is limited within a \((N_{\text{search}} \times N_{\text{search}})\) area around he feature point. The similarity measure for the match is the sum of absolute differences between the template and all the \(N_t \times N_t\) square sub-images within the search space:

\[
SAD(d_x, d_y) = \sum \left| \text{template}(x, y) - \text{images}(x + d_x, y + d_y) \right|
\]  

(1)

The objective function is:

\[
\arg\min_{d_x, d_y} \left( SAD(d_x, d_y) \right)
\]  

(2)

3.3. Three-dimensional reconstruction algorithm

In this three-dimensional reconstruction process, photography are taken for the same ground target point in the two stations \(C_1\) and \(C_2\) respectively, as shown in Figure 3, where \(C_1 X_{C_1} Y_{C_1} Z_{C_1}\) is the camera coordinate system in station \(C_1\), and \(C_2 X_{C_2} Y_{C_2} Z_{C_2}\) is the camera coordinate system in station \(C_2\). The image points of the target on the ground \(P\) on the photos (left and right) are \(p_1\) and \(p_2\). Obviously, the rays of \(C_1p_1\) and \(C_2p_2\) intersect at the target point on the ground \(P\).

![Figure 3. Cross target localization](image)

The collinear equation of \(C_1\) and \(C_2\) is:

\[
\begin{align*}
x_i &= F_x \frac{x_c}{z_c} + C_x \\
y_i &= F_y \frac{y_c}{z_c} + C_y
\end{align*}
\]  

(3)

where \((x_i, y_i), i=1,2\) are the pixel coordinates of \(P\). \((F_x, F_y)\) is the focal length of camera. \((C_x, C_y)\) is the image principal point coordinate, \((x_c, y_c, z_c), i=1,2\) are the coordinates of \(P\) in the camera coordinate system of each station.

Through the arrangement of equation (3):

\[
\begin{align*}
x_c &= \frac{(x_i - C_x) \cdot z_c}{F_x} \\
y_c &= \frac{(y_i - C_y) \cdot z_c}{F_y}
\end{align*}
\]  

(4)

During the flight, the attitude angle of UAV, the azimuth angle and elevation angle of the camera will change. The UAV attitude can be measured by airborne IMU; azimuth angle and elevation angle of the camera can be obtained by the camera pan and tilt angle encoder. Through the coordinate transformation, the coordinate of \(P\) in the geographic coordinate system of the stations \(C_i\) can be expressed as:

\[
\begin{bmatrix}
x_m \\
y_m \\
z_m
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \phi & -\sin \phi \\
0 & \sin \phi & \cos \phi
\end{bmatrix}
\begin{bmatrix}
\cos \gamma_c & 0 & \sin \gamma_c \\
0 & \cos \theta_c & -\sin \theta_c \\
-\sin \gamma_c & \cos \theta_c & \cos \alpha_c
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \beta_c & -\sin \beta_c \\
0 & \sin \beta_c & \cos \beta_c
\end{bmatrix}
\begin{bmatrix}
x_o \\
y_o \\
z_o
\end{bmatrix}
\]  

(5)
where \((x_i, y_i, z_i), i = 1, 2\) are the coordinates of \(P\) in the geographic coordinate system of the stations \(C_i\). \(\phi, \gamma, \theta\) are the aircraft heading angle, pitch angle and roll angle. \(\alpha_i, \beta_i\) are the camera azimuth angle and elevation angle.

Combined with (4) and (5),
\[
\begin{align*}
x_{N0} &= f_1(z_i, x_i, C_i, F_i, \phi_i, \gamma_i, \theta_i, \alpha_i, \beta_i), \\
y_{N0} &= f_2(z_i, y_i, C_i, F_i, \phi_i, \gamma_i, \theta_i, \alpha_i, \beta_i), \\
z_{N0} &= f_3(z_i, \phi_i, \gamma_i, \theta_i, \alpha_i, \beta_i)
\end{align*}
\]

(6)

Since the distance between the stations \(C_1\) and \(C_2\) is short, the effect of Earth curvature can be ignored. The geographic coordinate system of station \(C_2\) is only the translation of geographic coordinate system of station \(C_1\) without the rotation of coordinate system. For calculation convenience, the geographic coordinate system of station \(C_1\) is taken as the reference coordinate system, and the latitudes, the longitudes and the altitudes of stations \(C_1\) and \(C_2\) can be measured by the onboard GPS. Through coordinate transformation, the coordinate values of station \(C_2\) relative to the reference coordinate system can be calculated. We set the position of target \(P\) in the reference coordinate system as \((x_p, y_p, z_p)\),
\[
\begin{align*}
x_p &= x_{N1} + \Delta x, \\
y_p &= y_{N1} + \Delta y, \\
z_p &= z_{N1} + \Delta z
\end{align*}
\]

(7)

Combined with (6) and (7), image point coordinate of \(P\) \((x_i, y_i)\), the equivalent focal length of the camera \((F_x, F_y)\), and image principal point coordinate \((C_x, C_y)\) are all known; the parameters \(\phi, \gamma, \theta, \alpha_i, \beta_i, \Delta x, \Delta y, \Delta z\) can be measured; the variables to be solved are \(x_p, y_p, z_p, \Delta x, \Delta y, \Delta z\). The coordinate of target point \(P\) in the reference coordinate system can be solved through equation (7).

In the above method, the calculation results are very sensitive to various noises. This is because the distance between target and camera is much larger than the focal length. In order to improve the algorithm robustness and localization accuracy, we propose a improved algorithm, in which photography are taken for multiple times on the same target point, and then using least square iterative method to solve the optimal solution.

4. Experiment
We used our UAV experimental platform to obtain the experimental image data. The experimental area mainly includes farmland, trees, roads and other objects. The flight height is set to be 5m, the image size is 640 pixel \(\times\) 480 pixel, with a total of 89 images. The camera undergoes a rigorous calibration and the image distortion is corrected based on calibration parameters.

Figure 4. UAV platform and ground station software

In order to verify the performance of this algorithm, two groups of contrast experiments are carried out in the same platform (Intel Core i7/8G memory).
4.1. Algorithm efficiency analysis
In order to test the efficiency of the proposed algorithm, the processing time of feature points extraction, feature points matching as well as 3D reconstruction are statistically analyzed respectively, which are compared with those by the Bundler algorithm, as shown in Table 1. The efficiency of the proposed algorithm is significantly improved. The total time is reduced from 390.9s to 31.5s, and the efficiency is increased by at least 12.4 times.

4.2. Reconstruction effect analysis
The two algorithms can reconstruct the three-dimensional structure of the region. However, 54321 three-dimensional points are reconstructed by the proposed algorithm, while the Bundler only reconstructed 33115 points. The number of reconstructed points is 63.9% larger than that of Bundler. However, the uniformity of the point cloud constructed by the proposed algorithm is worse than that of the Bundler algorithm, because in order to guarantee the real-time performance, the algorithm directly adopts the UAV attitude data provided by the sensor to calculate, and the error of the sensor leads to the target point position error. The Bundler algorithm is to compute the re-projection error after obtaining the three-dimensional point, and eliminate the point in which the re-projection error is greater than a certain threshold value. The retained points, camera attitude and position parameters are combined with the adjustment, so as to improve the positioning accuracy of the target point.

Table 1. Performance comparison between proposed algorithm and bundler algorithm

|                  | Feature Points Extraction | Feature Points Matching | 3D Reconstruction | Total Time |
|------------------|---------------------------|-------------------------|-------------------|------------|
| Proposed Algorithm | 16.7s                     | 11.5s                   | 3.3s              | 31.5s      |
| Bundler          | 10.2s                     | 112.4s                  | 268.3s            | 390.9s     |

5. Conclusion
In this paper a novel method of farmland 3D terrain reconstruction based on UAV sequence images is proposed. We use SILC image segmentation algorithm and Harris corner detector algorithm to obtain feature points and use SAD algorithm to match feature points. A three-dimensional reconstruction algorithm according to monocular sequence images, UAV position, attitude and camera pointing angle is derived. The results of the experiments demonstrated that this method can quickly and effectively reconstruct farmland terrain, which can provide a data base for UAV high-fixing and has important application value.

Acknowledgements
State Key Laboratory of Satellite Navigation System and Equipment Technology (EX166840046), the National Natural Science Foundation of China (Grant No. 61601222), the Natural Science Foundation of JiangSu Province (Grant No. BK20160789) and China Postdoctoral Science Foundation Funded Project (Grant No.2018M632303).

References
[1] Wendel J, Meister O, Schlaile C and Trommer GF 2006 An integrated GPS/MEMS-IMU navigation system for an autonomous helicopter J. Aerospace Science and Technology. 10(6) 527-33
[2] Anthony D, Elbaum S, Lorenz A and Detweiler 2014 On crop height estimation with UAVs IEEE/RSJ International Conference on Intelligent Robots and Systems (Amsterdam: Elsevier) pp 4805-12
[3] bin Misnan M F, Arshad N M and Razak N A 2012 Construction sonar sensor model of low altitude field mapping sensors for application on a UAV 2012 IEEE 8th International Colloquium on Signal Processing & its Applications (USA: IEEE) pp 446-50
[4] Geiger A, Ziegler J and Stiller C 2011 Stereoscan: Dense 3d reconstruction in real-time
Intelligent Vehicles Symposium (IV), 2011 IEEE. (USA: IEEE) pp 963-68
[5] Snavely N, Seitz S M and Szeliski R 2006 Photo tourism: exploring photo collections in 3D
CACM transactions on graphics. 25(3) 835-46
[6] Achanta R, Shaji A, Smith K, Lucchi A, Fua P and Susstrunk S 2012 SLIC superpixels compared
to state-of-the-art superpixel methods J. IEEE transactions on pattern analysis and machine
intelligence. 34(11) 2274-82