SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

August 2024, Vol. 67, Iss. 8, 182305:1–182305:15
https://doi.org/10.1007/s11432-024-4056-8

An adaptive 3D reconstruction method for asymmetric dual-angle multispectral stereo imaging system on UAV platform

Chen WANG[1], Xian LI[1], Yanfeng GU[1*] & Zixu WANG[2]

[1]*School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin* 150001*, China;*
[2]*Beijing Institute of Control and Electronic Technology, Beijing* 100038*, China*

Received 8 January 2024/Revised 20 March 2024/Accepted 31 May 2024/Published online 25 July 2024

Abstract A multispectral imaging system often cannot capture 3D spatial information owing to hardware limitations, which diminishes the effectiveness across various domains. To address this problem, we have developed a multispectral stereo imaging system along with an adaptive 3D reconstruction algorithm. Unlike existing unmanned aerial vehicle stereo imaging systems, our multispectral stereo imaging system uses two multispectral cameras with asymmetric spectral bands positioned at different angles. This design enables the acquisition of a higher number of bands and lateral spatial information while maintaining a lightweight structure. This system introduces challenges such as large geometric distortions and intensity differences between multiple bands. To accurately recover 3D spatial information, we propose an adaptive 3D reconstruction method. This method employs a position and orientation system-assisted projection transformation and a normalized threshold adjustment strategy. Finally, mutual information is used to reconstruct the multispectral images densely, effectively addressing nonlinear differences and generating a comprehensive multispectral point cloud. Our stereo system was used for two real data collections in different regions, and the efficacy of the proposed 3D reconstruction method was validated by comparing it with existing methods and commercial software.

Keywords multispectral images, 3D reconstruction, stereo imaging system, unmanned aerial vehicle, feature extraction

1 Introduction

A multispectral imaging system equipped on an unmanned aerial vehicle (UAV) can flexibly and cost-effectively capture multispectral images, which contain rich spectral information of the feature surface [1–5]. Based on light-splitting devices, multispectral cameras can be divided into filter wheel and multi-lens types [6–8]. The former consists of a panchromatic camera and a set of filters that can acquire multispectral images with high spatial resolution. However, owing to the inability to image multiple bands at the same time, alignment errors between images of different bands are unavoidable. The latter offers a resolution to this problem by independently imaging each spectral band through its corresponding lens. Compared to RGB cameras, multispectral cameras provide additional spectral bands that offer more effective information for applications in agriculture, emergency response, fire monitoring, and other fields [9–11]. However, multispectral imaging systems cannot capture 3D spatial information owing to hardware limitations, thereby restricting their real application in various domains where crucial elevation information is needed [12].

Multispectral stereo imaging systems are developed to acquire multispectral images from several different viewpoints for stereo imaging. These systems can be classified based on their deployment platform (ground-based or airborne) and the number of cameras used (one, two, or more). Ground-based equipment typically includes vehicle-mounted and handheld devices. Smith [13] developed a multispectral stereo imaging system for the Mars Pathfinder Imager. This system features two multispectral cameras placed in parallel to detect terrain and rock information at a distance of 20–30 m. Kazemzadeh et al. [14]

* Corresponding author (email: guyf@hit.edu.cn)

designed a handheld multispectral stereo imaging device that captured multispectral images from three angles and nine bands using three lenses and a set of mirrors tilted on the imaging axis.

Although ground-based platforms have better imaging stability, their acquisition range and efficiency are far lower than those of airborne platforms. Leveraging the movement capabilities of UAV platforms, even a single multispectral camera can capture images from different viewpoints for 3D reconstruction [15]. Zainuddin et al. [16] implemented a multi-camera multispectral imaging system on a UAV to facilitate 3D reconstruction of petroglyphs. They employed a single-angle multispectral camera for stereoscopic imaging, which is primarily suitable for scenes with minimal surface undulations. Stereo imaging systems usually consist of two or more cameras with different viewpoints to obtain more comprehensive stereo information [17]. To increase the angle of multispectral image acquisition, Briechle et al. [18] designed a stereo imaging system featuring two identical multispectral cameras in a twisted configuration. However, the inclusion of two identical multispectral cameras in the stereo imaging system resulted in increased weight without a corresponding increase in the number of wavelength bands.

Beyond the hardware components of the multispectral stereo imaging system, developing a multispectral 3D reconstruction algorithm is crucial. This algorithm falls under the category of image-based 3D reconstruction algorithms, which have become a significant research focus. These techniques can recover the 3D spatial information of the observed scene from RGB image sets [19–21]. Extending these methods to multispectral images can yield 3D data enriched with additional spectral information, expanding the range of application scenarios, such as agriculture [22, 23]. Image-based 3D reconstruction involves reconstructing a 3D model of an object from multiple 2D images, an important research direction in computer vision [24, 25]. 3D reconstruction methods usually require several steps, including camera parameter estimation, sparse reconstruction, and dense reconstruction. Camera parameter estimation is key to determining the reconstruction accuracy, involving the calculation of internal and external parameters. The internal parameters need to calibrate the camera to correct the camera's focal length, principal point position, and distortion coefficient, while the external parameters need to compute and optimize the camera's relative position and attitude through feature extraction, feature matching, and bundle adjustment. Owing to significant variations in viewing angles and irregular image overlap in UAV-collected images, robust feature extraction methods are critical [26]. To satisfy the computational cost requirements of 3D reconstruction, feature extraction methods need to be highly efficient, especially when dealing with large aerial imagery data sets [27–29]. Sparse reconstruction uses the optimized pose relationship and featured point locations to employ triangulation principles, obtaining the 3D spatial information of feature points. Some RGB-based reconstruction methods, such as COLMAP [30], have achieved notable results. In addition, commercially available reconstruction software has seen widespread adoption.

Existing 3D reconstruction algorithms primarily focus on RGB images, with limited research conducted on reconstruction algorithms for multispectral images. Algorithms for multispectral stereo imaging are divided into two categories based on the number of input images: single-shot stereo imaging and 3D reconstruction from image sets. Single-shot stereo imaging applies to systems with two cameras fixed on the same plane without considering the camera's positional information [31, 32]. Stereo information is obtained utilizing depth map estimation methods. Image-based 3D reconstruction requires the estimation of external parameters, and the majority of existing multispectral 3D reconstruction studies treat multispectral images as monochromatic images or convert them to panchromatic images [33, 34]. Recognizing the unique characteristics of multispectral images, Wang et al. [35] proposed a robust 3D reconstruction method specifically for multispectral images.

Despite some progress, there are still limitations in the research of multispectral stereo imaging systems and their corresponding reconstruction algorithms. Multispectral imagers are also typically larger than RGB cameras owing to their imaging mode. Multispectral stereo cameras for UAVs face mutual constraints related to the number of bands, angles, and the weight of the equipment [36]. Current systems lack the capability to simultaneously acquire multispectral images from multiple angles and bands. In terms of reconstruction algorithms, the multiview reconstruction of multispectral images faces some challenges, such as geometric distortions caused by multiple viewpoints and nonlinear radiometric differences between different bands. These issues significantly impact the precision and completeness of 3D reconstruction for multispectral images.

To achieve a more lightweight multispectral 3D reconstruction with an increased number of bands, we have developed a dual-angle asymmetric multispectral stereo imaging system comprising two distinct multispectral cameras operating in different spectral ranges. The angle between the main optical axes

of the dual cameras is set to 60° to acquire multispectral images from different angles. The system can acquire images from more angles and more spectral bands of the scene from the hardware level, but it also introduces challenges to 3D reconstruction algorithms, specifically geometric distortion and spectral differences between the images acquired by the dual cameras. In this paper, we propose a dual-angle multispectral 3D reconstruction method for the multispectral stereo imaging system. First, the stereo multispectral imaging system is calibrated utilizing a dual checkerboard plane to provide accurate position and orientation system (POS) information for the two multispectral cameras. With the assistance of initial positional information, multispectral images from different angles are projection-transformed to minimize geometric distortions. We then introduce a threshold adaptive adjustment method based on normalized cross-correlation (NCC) for feature extraction of cross-band image pairs. Finally, multiview stereo (MVS) techniques based on mutual information (MI) are used to acquire multiband dense reconstruction results. To assess the effectiveness of the proposed method, we conducted data acquisition and reconstruction experiments using a dual-angle multispectral stereo imaging system carried by UAV. Experimental results from two regions show that the proposed method can obtain satisfactory multi-band reconstruction results. The main contributions of this study can be summarized as follows.

(1) A 3D reconstruction workflow applicable to a dual-angle asymmetric multispectral stereo camera is proposed to successfully generate multispectral point cloud (MSPC) data with more bands. To the best of our knowledge, this study represents the pioneering use of a UAV platform equipped with a multispectral stereo system capturing data from various angles and bands.

(2) To address the intensity differences caused by the asymmetry of the dual-camera bands, we developed a cross-band adaptive feature extraction algorithm. This algorithm dynamically adjusts the threshold value based on normalized correlation coefficients between images from different bands, thereby increasing the number of corresponding feature points across bands and enhancing the robustness of multispectral 3D reconstruction.

(3) To address the geometric distortion caused by dual-angle imaging, this paper proposes an image matching method that elegantly combines projection transformation and block re-matching. Projection transformation enables geometric correction for images captured from different angles, thereby reducing the impact of geometric differences in feature matching. The block re-matching algorithm leverages POS information to minimize matching errors and enhance reconstruction accuracy.

The rest of the paper is organized as follows. Section 2 describes in detail the dual-angle multispectral stereo imaging system and the proposed 3D reconstruction method. Section 3 presents the experimental results and analysis of the proposed method on two datasets. The conclusion is drawn in Section 4.

2 Multispectral stereo imaging system and reconstruction method

2.1 Asymmetric multispectral stereo imaging system

The asymmetric multispectral stereo imaging system is illustrated in Figure 1(a). The key point of this system is the design of the camera bands and imaging angles. For band selection, we chose the camera band of Rededge-MX dual, which covers all bands from visible to near-infrared. This camera set consists of two multispectral cameras with distinct band configurations, as illustrated in Figure 1(b). The red multispectral camera has a wavelength range of 459–870 nm, and the blue camera has a wavelength range of 430–749 nm. The availability of ten different spectral ranges provides abundant opportunities for diverse applications in forestry, agriculture, and other relevant fields. In terms of angle design, the inclination angle of commonly used oblique photogrammetry cameras typically ranges from 15° to 50°. Given the substantial weight of multispectral cameras, nadir cameras are not included in the stereo imaging system of this UAV. Therefore, a 30° inclination angle was selected for the multispectral stereo imaging system, considering the acquisition of top and side information to enhance the robustness of stereo imaging. Compared to a single nadir camera, the dual-angle imaging system can capture more information about the observed scene on the UAV platform and reduce the missing sides in the reconstruction results (as shown in Figures 1(c) and (d)).

The designed multispectral stereo imaging system consists of two cameras and a downwelling light sensor module (DLS) with the POS. Both cameras possess a resolution of 1280 × 960, a field of view of 47.2° and exhibit a combined weight of 508.8 g. Two multispectral cameras were fixed in a rigid connection. The two cameras were opposed to each other, each at an angle of approximately 30° to

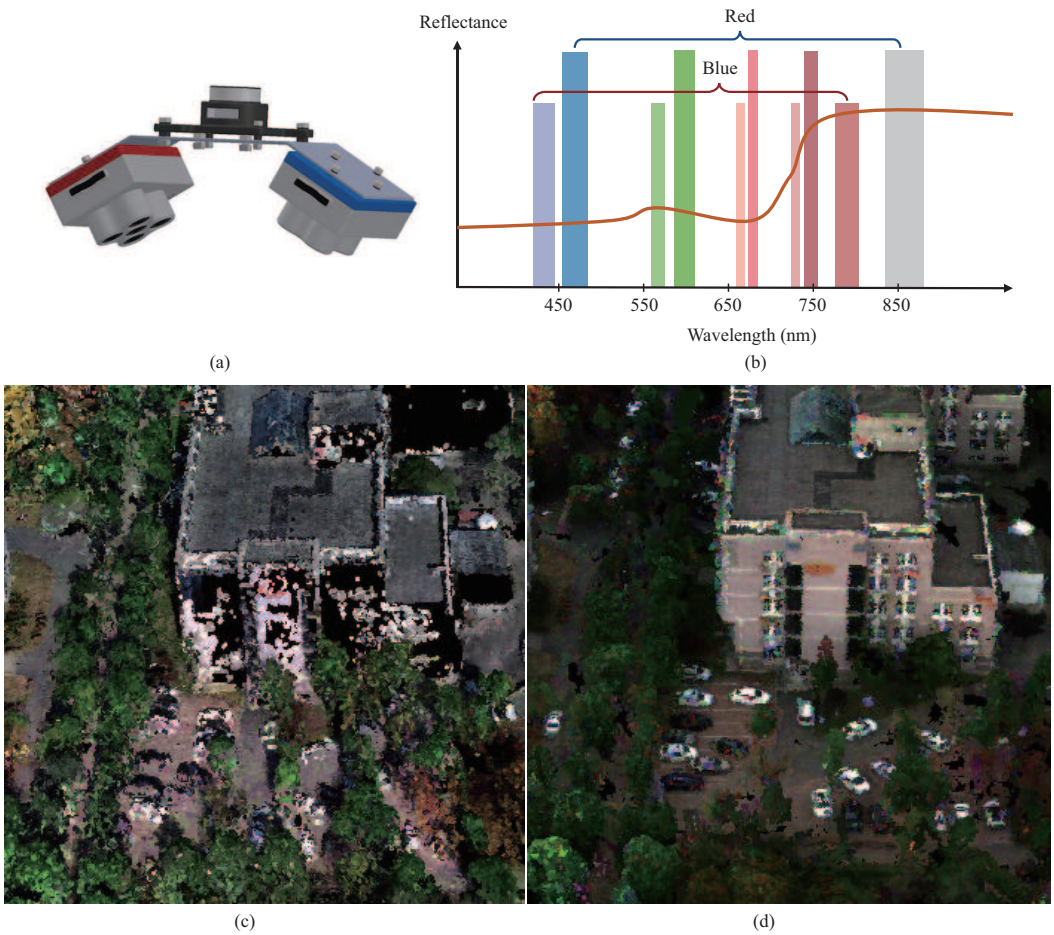


Figure 1 (Color online) Multispectral cameras and stereo imaging system. (a) Stereo multispectral imaging system consisting of multispectral cameras; (b) illustration of the bands of the multispectral imaging system; (c) reconstruction result of the nadir system; (d) reconstruction result of our system.

the vertical plane. The dual-angle stereo imaging system is capable of acquiring spectral information from more sides, and the two multispectral cameras with different wavelength bands are complementary to each other. Cameras communicate over a wired connection for synchronized imaging. Through the design of asymmetric spectral bands, our system further expands the spectral band number based on the simultaneous collection of dual-angle images, and realizes the lightweight design to adapt to the UAV platform. However, due to the relationship between field of view and imaging angles, our system is unable to capture images with overlapping fields through a single imaging process when mounted on the UAV. Therefore, it is imperative to propose a multispectral stereo reconstruction method for multispectral images acquired from diverse locations.

2.2 Adaptive multispectral stereo reconstruction method

According to the designed dual-angle multispectral stereo imaging system, the intensity differences and geometric distortions between the cameras pose challenges to the 3D reconstruction. An adaptive multispectral stereo reconstruction method is proposed, which mainly includes three steps: non-overlapping field dual-camera calibration, spectral-angle adaptive multispectral feature extraction and matching, sparse reconstruction and multi-dense matching based on MI. Sparse reconstruction based on structure from motion (SFM) [37] and dense reconstruction of multispectral multi-view stereo (MVS) [38] are used in the reconstruction process. The flowchart of the 3D reconstruction of multispectral images is shown in Figure 2, where the input is a multiband multispectral image set, and the output is a densely reconstructed MSPC. Details of these three components are described below.

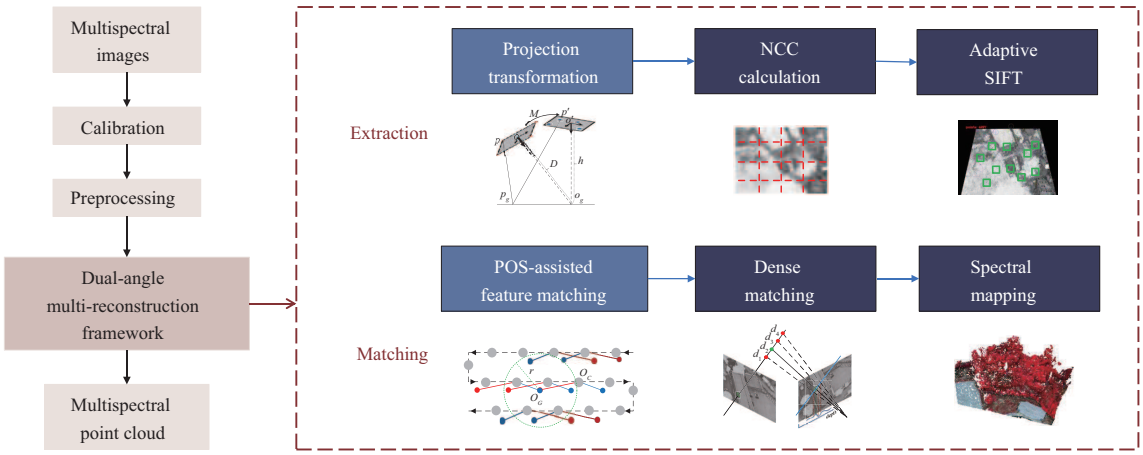


Figure 2 (Color online) Flowchart of the proposed dual-angle multispectral reconstruction. In the dual-angle multispectral reconstruction framework, the light box part is dedicated to solving the geometric challenges caused by the angles, and the dark box part is the module that focuses on the spectral difference between the bands.

2.2.1 *Non-overlapping field dual-camera calibration*

The DLS module contains POS in addition to the ambient light sensor. The global navigation satellite system in POS was designed to be placed above the UAV and was shared by both cameras. The inertial measurement unit is employed to record attitude angle (yaw, pitch and roll) data at the moment of camera exposure, with the cameras parallel to the sensor by default. Therefore, after adjusting the camera placement angle, the imaging plane of the camera is different from the DLS plane. The dual-angle cameras require calibration to correct the attitude information obtained from the DLS module.

The calibration parameters of the camera include internal and external parameters, where the internal parameters include the focal length, the principal point and the distortion coefficients. The external parameters include the relative vectors and the rotation matrix between cameras. Typically, the checkerboard calibration method can be used to calibrate a single camera. However, in the developed dual-angle stereo imaging system, there is a large difference in viewing angle between the cameras, which makes it impossible to capture a complete image of the same checkerboard in a single imaging session. A non-overlapping field dual-camera calibration method is used in this method to estimate the internal and external camera parameters. The calibration process is shown in Figure 3. The calibration process is estimated by building equations for multiple sets of checkerboard corner points and their corresponding image pixel points as follows:

$$\min \sum_{i=1}^{q} \sum_{j=1}^{p} \left\| \boldsymbol{m}_{ij} - \boldsymbol{m}'(\boldsymbol{K}, \boldsymbol{R}_i, \boldsymbol{t}_i, \boldsymbol{X}_{ij}) \right\|^2, \tag{1}$$

where q is the number of images, p is the number of corner points, $\boldsymbol{m}'(\boldsymbol{K}, \boldsymbol{R}_i, \boldsymbol{t}_i, \boldsymbol{X}_{ij})$ represents the coordinates of the image plane projection of the jth checkerboard intersection point of the ith image, and $\boldsymbol{m}_{ij}$ is the coordinates of the point on the image. The proposed method utilizes a nonlinear optimization algorithm to estimate the internal reference matrix $\boldsymbol{K}$ and the external parameters $\boldsymbol{R}_i, \boldsymbol{t}_i$.

The external parameters of one camera are transformed to the reference coordinate system of the other camera to realize the solution of the relative external parameters of the two cameras.

$$\boldsymbol{p}_1' = \begin{bmatrix} \boldsymbol{R}_c & \boldsymbol{t}_c \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \times \boldsymbol{p}_1, \tag{2}$$

$$\boldsymbol{p}_1^2 = \begin{bmatrix} \boldsymbol{R}_2 & \boldsymbol{t}_2 \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \times \begin{bmatrix} P_2^X \\ P_2^Y \\ P_2^Z \\ 1 \end{bmatrix} - \boldsymbol{R}_2 \times \boldsymbol{t}, \tag{3}$$

where $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are a pair of points on the two checkerboard calibration boards. $\boldsymbol{p}_1'$ is the transformed camera C_1 coordinate of point $\boldsymbol{p}_1$, which should be consistent with $\boldsymbol{p}_1^2$ (the coordinate of point $\boldsymbol{p}_1$ in the

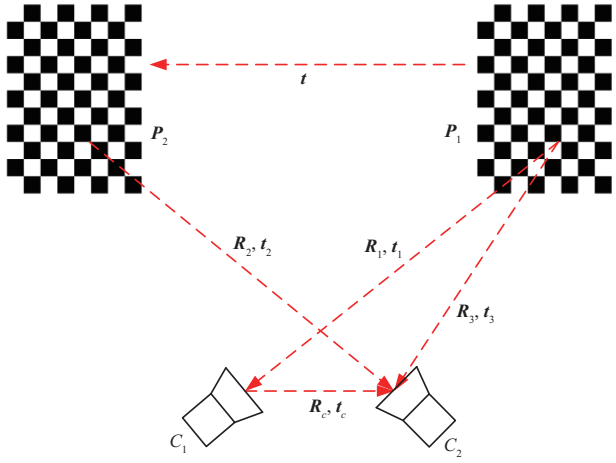


Figure 3 (Color online) Calibration of the dual-angle multispectral stereo imaging system using double checkerboard panels.

camera C_2 coordinate system).

$$\left[\boldsymbol{R}_2\ \boldsymbol{t}_2 - \boldsymbol{R}_2 \times \boldsymbol{t}\right] = \left[\boldsymbol{R}_c\ \boldsymbol{t}_c\right] \times \begin{bmatrix} \boldsymbol{R}_1 & \boldsymbol{t}_1 \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}. \tag{4}$$

According to the corresponding relationship, we can derive (4) and solve it. The obtained rotation matrix, $\boldsymbol{R}_c$, and offset vector, $\boldsymbol{t}_c$, represent a set of relative external parameter calibration results for the dual cameras. The solution can be optimized with multiple pairs of points in images to obtain the final solution with minimum reprojection error.

2.2.2 *Spectral-angle adaptive multispectral feature extraction and matching*

Feature point extraction and matching are key steps in determining the accuracy of 3D reconstruction results. Common feature extraction methods can be categorized into region-based and feature-based [39]. Region-based methods are not suitable for the alignment of images with large rotational or perspective transformations, while geometric feature-based methods are usually ineffective for matching images with large radiometric variations [40]. Therefore, a threshold adaptive adjustment cross-band feature extraction method is used in this step, and the extraction step includes POS-assisted perspective projection transformation, normalized correlation coefficients (NCC) calculation, scale invariant feature transform (SIFT) feature extraction and projection inverse transformation of feature points.

First, reflectance correction and band alignment are performed on the input image, which is used to obtain accurate spectral values [35]. Dual angle oblique imaging leads to more severe geometric distortions in the image compared to single angle camera imaging. The same object is significantly different in images from different angles. Based on the camera imaging model, a perspective transformation is performed on the tilted image to minimize geometric distortion with the assistance of the checked POS information. The oblique images of different angles are roughly orthorectified to unify the geometric coordinates. The projection model is as follows:

$$\boldsymbol{P}' = \boldsymbol{M}\boldsymbol{p}, \tag{5}$$

$$\boldsymbol{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}, \tag{6}$$

$$x' = \frac{X}{Z} = \frac{m_{11}x + m_{12}y + m_{13}}{m_{31}x + m_{32}y + m_{33}}, \tag{7}$$

$$y' = \frac{Y}{Z} = \frac{m_{21}x + m_{22}y + m_{23}}{m_{31}x + m_{32}y + m_{33}}, \tag{8}$$

where $\boldsymbol{p}(x, y, 1)$ is the image coordinates before projection, $\boldsymbol{P'}(X, Y, Z)$ is the spatial coordinates. $\boldsymbol{M}$ represents the perspective projection matrix. When the altitude of the UAV is known, it can be solved using the principle of perspective projection through the internal and external parameters of the camera.

The calibrated image reduces most of the geometric distortions. However, there are still differences in radiation intensity between images of different bands. A combination of region-based and geometric features was used for feature extraction. The SIFT operator, as a classical feature extraction operator, builds a pyramid scale space by continuously reducing the size of the image and applying a Gaussian kernel for convolution, and then finds the feature points with scale invariance on different scale spaces. Subsequently, the orientation and high dimensional descriptor vectors of the feature points are computed. The algorithm is robust to scale, rotation and illumination and is more robust in various complex images. Extreme point detection is performed on difference Gaussian pyramid $\boldsymbol{D}(x, y, \sigma)$ as follows:

$$\boldsymbol{D}(x, y, \sigma) = \boldsymbol{L}(x, y, k\sigma) - \boldsymbol{L}(x, y, \sigma), \tag{9}$$

$$F_{\mathrm{MAX}}(x, y) > cT, \tag{10}$$

where $\boldsymbol{L}(x, y, \sigma)$ is the original image convolved with the Gaussian blur at scale σ. $F_{\mathrm{MAX}}(x, y)$ is the intensity value of the potential extreme point. cT is the contrast threshold for extraction, which is used to filter noise.

Due to the intensity differences between different bands, it is difficult to determine a uniform threshold when applied to multispectral images as in (10). Therefore, the NCC is calculated for the images in different bands after projection as (11), and the correlation coefficients represent the degree of similarity of the images [41, 42].

$$\mathrm{NCC}(i, j) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\left[R^{i,j}(m, n) - \overline{R^{i,j}}\right] \times \left[T^{m,n} - T\right]}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}\left[T^{i,j}(m, n) - \overline{T^{i,j}}\right]^2 \cdot \sum_{m=1}^{M}\sum_{n=1}^{N}\left[T^{m,n} - T\right]^2}}, \tag{11}$$

where R is the reference image, T represents the template image. (i, j) is the coordinate of the reference point.

A larger similarity represents a high number of potentially identical matching features. The number of feature extraction points was set according to the size of the correlation coefficient, and the feature extraction threshold was adaptively decreased from large to small until the number of feature extraction requirements was met.

In the feature matching stage, since the large number of images in the large scene 3D reconstruction task, it is extremely inefficient to match all the images. Each image covers a small ground area, and there are only a few overlapping areas between images. Therefore, a matching image pair search method is utilized to reduce the number of images matches significantly by leveraging the camera's initial positional data. The coordinates of the projected position of the image can be calculated by (7) and (8).

In addition, the block re-matching algorithm is employed to improve the matching accuracy of feature points. The initial matching results are utilized to estimate the homography transformation relationship between image pairs. According to the transformation matrix, the image feature points to be matched are projected onto a reference image as Figure 4(a). The reference image is segmented and local re-matching is performed within the block. This method can reduce the cases of mis-matching and missing matching during the global feature matching process, as shown in Figure 4(b).

2.2.3 *Sparse reconstruction and multi-dense matching based on MI*

SFM method is utilized for sparse point cloud generation based on the camera's internal references as well as feature matching relationships. The steps of the incremental SFM method include: restoring the relative pose relationship of the two initial images through epipolar geometric constraints, triangulating the feature points to solve for the 3D spatial coordinates of the feature points, estimating the newly added image pose through the PnP method and passing beam adjustment method is used for global optimization.

In the sparse reconstruction stage of dual-angle multispectral photogrammetry, images from the same angle tend to have more matching points, while images from different angles have fewer matching points, therefore we adopt different reconstruction criteria for images from different angles. In order to allow more images to participate in sparse reconstruction, and to avoid the existence of two independent

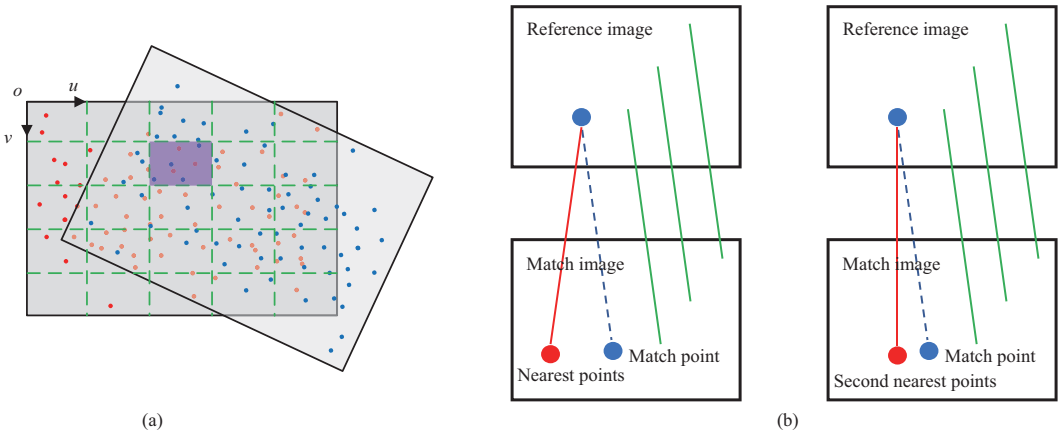


Figure 4 (Color online) POS-assisted feature matching. (a) Block re-matching. The red points are feature points in the reference image, and the blue points are feature points in the template image. (b) Diagrams of mismatch and missing match.

results from cameras at different angles in the reconstruction results, causing an increase in errors in global optimization. The adjusted reconstruction strategy includes the minimum required number of reconstruction points, and the rules for priority matching. For cameras from different angles, we reduce the requirement on the number of matching pairs and select an appropriate threshold based on the size of the reconstruction error. In terms of matching rules, according to the distance of the backward projection point, a reconstructed image with a moderate baseline length is selected to improve the reconstruction accuracy.

In terms of dense matching, the image's external parameters and the camera's internal parameters are used to estimate the depth map of each image. In a set of matching image pairs, for each pixel on the reference image, the epipolar line on the matching image can be obtained through epipolar geometry. Then the depth of each pixel is obtained through consistency estimation. In order to reduce the inaccuracy in depth estimation caused by the intensity difference between images in different bands, MI, which is suitable for non-linear intensity differences, is used as the consistency estimation criterion [43, 44].

$$\mathrm{MI}(R,T) = H(R) + H(T) - H(R,T), \tag{12}$$

$$H(R) = -\sum_{r\in R} P(r)\log_2 P(r), \tag{13}$$

$$H(T) = -\sum_{t\in T} P(t)\log_2 P(t), \tag{14}$$

$$H(R,T) = -\sum_{r\in R, t\in T} P(r,t)\log_2 P(r,t), \tag{15}$$

where $H(R)$ and $H(T)$ represent the entropy of the reference image block R and the template image block T respectively. $H(R,T)$ represents their joint entropy. The pixels r and t belong to image R and image T.

MI is a measure of statistics between two variables that is robust to nonlinear intensity differences. The results of the final dense reconstruction were calculated as in (13).

$$\boldsymbol{P}_i = \boldsymbol{R}^{\mathrm{T}}\left[D_i \times \boldsymbol{K}^{-1}\boldsymbol{p}_i - \boldsymbol{t}\right], \tag{16}$$

where D_i is the depth of point i, P_i is the coordinates of reconstruction point i.

3 Experiments and analysis

3.1 Data description

The experimental data were obtained by the multispectral stereo imaging system equipped on the DJI M300 UAV platform. The datasets include two different scenes: Harbin Institute of Technology (HIT)

(a) (b)

Figure 5 (Color online) Satellite images of the experimental scenes. (a) HIT campus. (b) ZJK Mangrove.

campus and Zhangjiangkou (ZJK) Mangrove (as shown in Figure 5). Dual-angle multispectral images and nadir view multispectral images were collected for experimental comparison.

(1) **HIT campus.** The scene contains several types of trees and buildings. The flight altitude of the UAV is 80 m and the flight speed is 8 m/s. The dataset contains 288 multispectral images for oblique imaging and 480 multispectral images for nadir view. The multispectral system employs a continuous shooting mode at consistent time intervals. To satisfy the image overlap rate requirements for three-dimensional reconstruction, a reasonable photo interval and flight strip spacing were chosen. The heading overlap rate of the images in the data set is 80%, and the side overlap rate is 60%.

(2) **ZJK Mangrove.** The scene contains artificial buildings and dense mangrove forests. The UAV flew at an altitude of 80 m and at a speed of 8 m/s. The dataset contains 248 multispectral images for oblique imaging and 240 multispectral images in nadir view. The multispectral system employs a continuous shooting mode at consistent time intervals. The overlap rate is 80% in the heading direction and 60% in the side direction.

3.2 Experimental setting

The experimental results are visually analyzed and quantitatively evaluated to illustrate the effect of each step. The quantitative evaluation indicators include: reconstructed image proportion, average number of feature extractions, reconstructed trajectory length, number of reconstructed features and reprojection error. In detail, the reconstructed percentage refers to the proportion of images in which the spatial relationships can be accurately reconstructed among all collected images. Feature number is the average of the features extracted from images. Track length is the number of images in the reconstructed tracks. A higher track length indicates better performance in feature extraction and matching during the reconstruction process. Reprojection error is defined as the projection error between reconstructed points and corresponding feature points, representing the accuracy of 3D reconstruction. Reconstruction feature refers to the number of reconstructed feature points. These indicators can evaluate the quality of 3D reconstruction from multiple aspects. Running times are also listed to illustrate reconstruction efficiency. To demonstrate the advantages of the proposed method, the experiments are compared with a variety of commonly used feature extraction methods in 3D reconstruction. The experimental results are compared with the multispectral image reconstruction method MODM [34], which processes multispectral images band by band. The experiments are also compared with commercial software Pix4d and well-known open source 3D reconstruction framework Colmap [25]. In addition, to demonstrate the advantages of dual-angle imaging, the reconstruction results of dual-angle multispectral are evaluated against the reconstruction results of multispectral images in nadir view [36].

3.3 Experimental results and analysis

The reconstruction results of the dataset are shown in Figures 6 and 7. The experimental analysis of the specific steps is as follows.

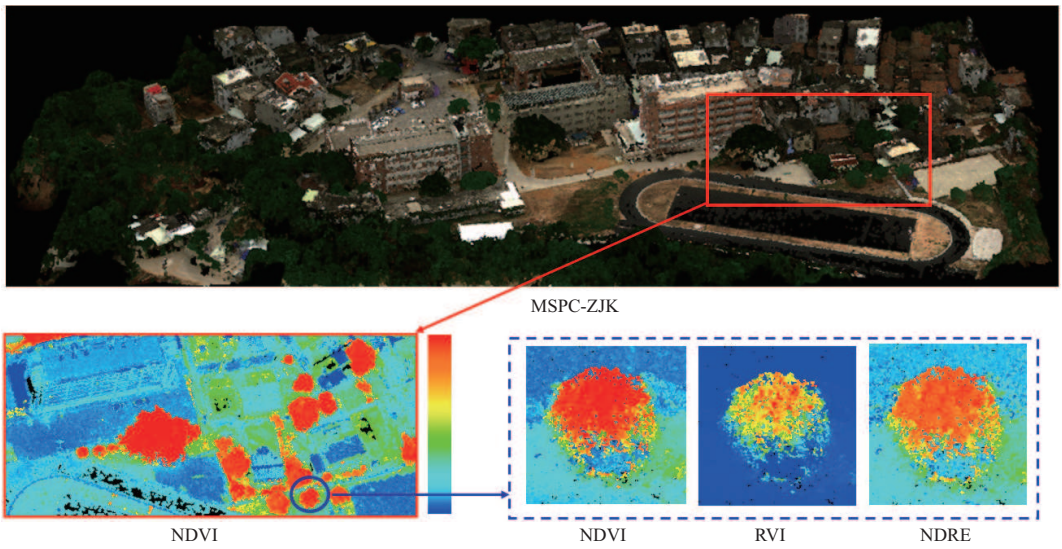


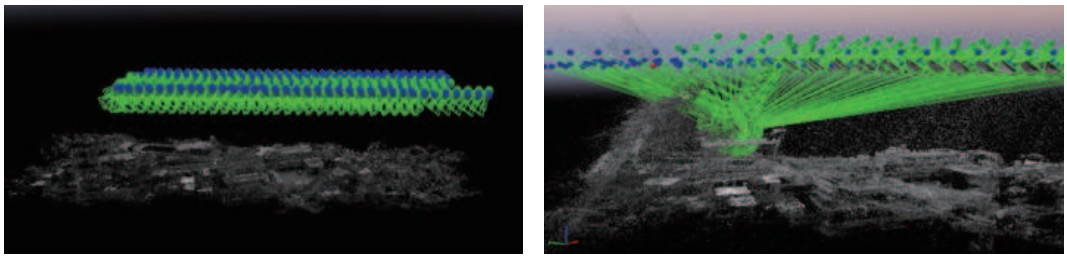
Figure 6 (Color online) Reconstruction results of ZJK. The upper part is the display of the RGB band. The lower part is the partial NDVI display and the rendering result of the vegetation index (NDVI, RVI, NDRE) of a tree.

Figure 7 (Color online) Reconstruction results of HIT. The right part shows the comparison of the local results of the contrast method MODM and the reconstruction results of nadir imaging.

Table 1 Cameras calibration results

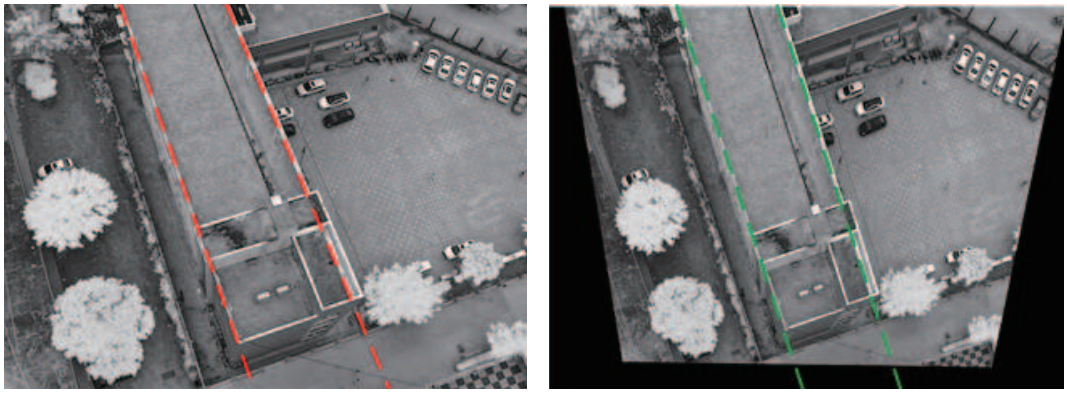
Parameters	Rededge MX-red	Rededge MX-blue
Equivalent focal length (f_x, f_y)	(1447.4, 1446.7)	(1459.5, 1456.7)
Principal point position (c_x, c_y)	(643.5, 489.1)	(647.2, 503.3)
Rotation matrix R_c	$\begin{bmatrix} 1.0000 & -0.0058 & 0.0036 \\ 0.0059 & 0.4799 & -0.8773 \\ 0.0034 & 0.8773 & 0.4800 \end{bmatrix}$	
Euler angles (°)	$(0.3386, -0.1921, 61.3176)$	
offset vector t_c	$[-5.0991, 122.8368, 64.5710]^{\mathrm{T}}$	

The calibration results of the dual-angle stereo imaging system are shown in Table 1. The calibration results are consistent with the designed camera's imaging angle. We corrected the POS data of the images collected by the two multispectral cameras according to the calibration parameters. The raw image and the processed image were subjected to reconstruction experiments respectively and the results are shown in Figure 8. When reconstructing the original image, errors in the initial positional relations cause distortions in the reconstruction results. The processed image obviously fixed this problem and the



(a) (b)

Figure 8 (Color online) Reconstruction results (a) before and (b) after calibration.



(a) (b)

Figure 9 (Color online) Oblique multispectral images (a) before and (b) after projective transformation.

correct orientation of images is displayed. For comparison of the contribution of the subsequent steps, all experiments were performed on the processed images.

In the image feature extraction stage, the projection transformation and adaptive thresholding feature extraction methods are proposed for geometric distortion and spectral intensity differences. Figure 9 illustrates the results of perspective projection of oblique images acquired by the multispectral cameras. As a result of the perspective effect, there is a clear non-parallel relationship between the two sides of the building in the image in the tilted image. In the ortho-corrected image, the parallel geometric relationship is restored. The geometric distortion between the images acquired by the cameras at different angles is reduced. It is also demonstrated that the feature extraction method containing the projection transform improves the ratio of the reconstructed image and reduces the reprojection error. In terms of intensity differences, the SIFT operator is robust to geometric differences such as scale and rotation, but is strongly influenced by the feature extraction threshold. The results of extracting features with a fixed threshold set and a fixed number of features extracted are shown in Figure 10. There is a significant difference in the number of features extracted in different bands at the same threshold. The strategy of a fixed number of features is to decrease the threshold for feature extraction until the number of extracted features surpasses the preset value. While extracting the same number of features from various bands, the disparities in band intensity lead to noticeable differences in feature distribution. The inhomogeneity of feature extraction caused by intensity differences leads to a reduction in the number of feature matching pairs. We aim to obtain more homonymous feature points from multispectral images in different bands acquired by two multispectral cameras for potential 3D reconstruction points.

Therefore, the NCC between the respective five-band images of the two cameras and the average number of feature matches under a fixed number of feature extraction strategies are calculated. Tables 2(a) and (b) list the NCC between different bands of the two data sets. The average number of matching feature pairs is shown in Tables 2(c) and (d). Based on each band of the blue camera, the maximum and second maximum values in the table are bolded. By comparing Tables 2(a)–(d), it can be concluded that the average number of feature matches is positively correlated with the band NCC. Based on this relationship, the local correlation coefficient in the projection transformed image is computed for the feature extraction thresholding decision. Different regions of the image are subjected to feature extraction using different thresholds based on the local NCC. More feature points are extracted from areas with greater correlation

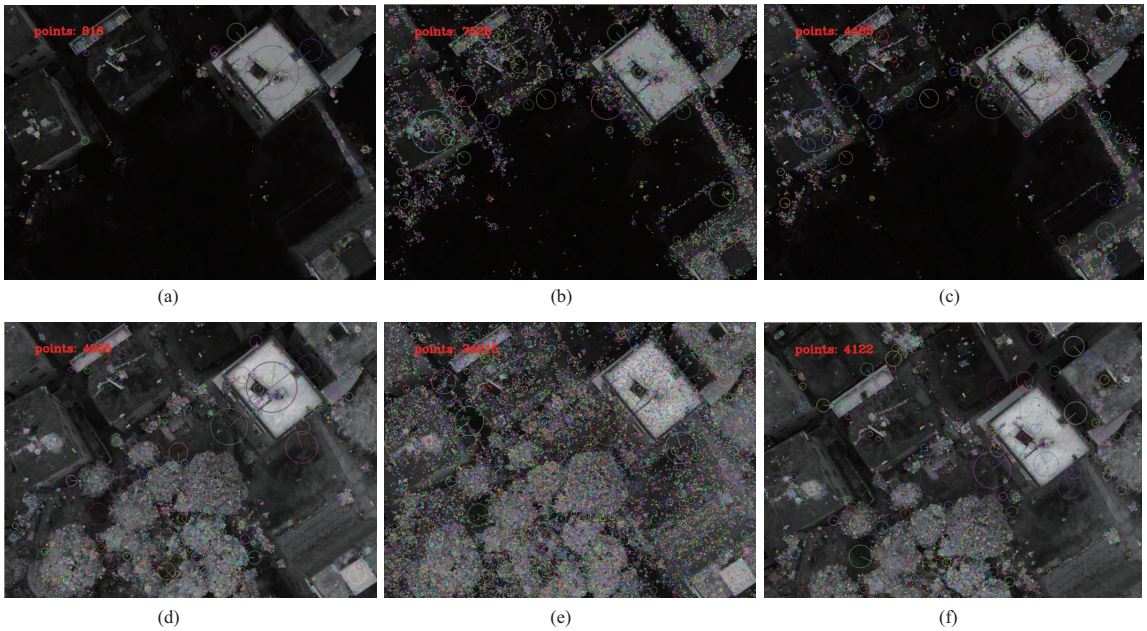


Figure 10 (Color online) Results of multispectral image feature extraction with different strategies. (a) The result of feature extraction with a fixed threshold in 531 nm band image; (b) the result with a fixed number of feature extractions in 531 nm band image; (c) the result of feature extraction with adaptive thresholding in 531 nm band image; (d) the result of feature extraction with a fixed threshold in 842 nm band image; (e) the result with a fixed number of feature extractions in 842 nm band image; (f) the result of feature extraction with adaptive thresholding in 842 nm band image.

Table 2 NCC and number of matched pairs between images of different bands[a)]

	(a) NCC between different bands in HIT					(b) NCC between different bands in ZJK				
	Blue = 444	Blue = 531	Blue = 650	Blue = 705	Blue = 740	Blue = 444	Blue = 531	Blue = 650	Blue = 705	Blue = 740
Red = 475	**0.61**	0.57	**0.58**	0.35	0.31	**0.72**	0.72	**0.74**	0.61	0.72
Red = 560	0.43	**0.48**	**0.49**	0.37	0.33	0.63	**0.66**	**0.69**	0.60	0.63
Red = 668	0.44	**0.47**	**0.50**	0.44	0.26	0.64	0.66	**0.69**	0.67	0.64
Red = 717	0.35	0.45	**0.68**	0.43	**0.66**	0.22	**0.28**	0.25	**0.29**	0.22
Red = 842	0.22	0.35	0.34	**0.38**	**0.58**	0.60	0.61	**0.83**	0.61	**0.64**
	(c) Average number of matching features in HIT					(d) Average number of matching features in ZJK				
	Blue = 444	Blue = 531	Blue = 650	Blue = 705	Blue = 740	Blue = 444	Blue = 531	Blue = 650	Blue = 705	Blue = 740
Red = 475	**236**	**202**	180	137	141	**350**	**336**	309	261	175
Red = 560	162	**253**	198	**206**	141	307	**350**	**355**	315	225
Red = 668	170	188	**320**	**205**	147	293	315	**431**	**370**	253
Red = 717	134	201	**213**	**281**	205	233	252	**314**	**315**	240
Red = 842	139	166	187	**206**	**258**	204	200	**236**	215	**235**

a) Blue and Red represent the blue camera and the red camera, the unit is nm.

to increase the number of matching feature points. According to the local NCC, different thresholds are applied to each part of the image for feature extraction. The quantitative evaluation metrics for the feature extraction stage are provided in Table 3. Compared to commonly used feature extraction methods, the features extracted by this method between different bands have achieved better results in feature matching.

In the feature matching stage, feature matching based on projection transformation reduces the number of images to be matched and further improves the computational efficiency in the feature matching stage. To improve the reconstruction accuracy, the block re-matching algorithm based on the projection relation further increases the number of features to be matched in the image set, but it also brings about an acceptable increase in computational cost. The running times of these methods are listed in Table 4.

The reconstruction strategy in 3D reconstruction techniques typically imposes no restrictions on camera angles for better robustness. In contrast, for dual-angle multispectral cameras, the strategy is modified to incorporate a larger number of images in the reconstruction process. The proposed method further optimizes the reconstruction strategy for dual-angle imaging systems, increasing the proportion of images

Table 3 Quantitative evaluation results of 3D reconstruction by different methods[a)]

Method	Reconstructed percentage (%)		Feature number		Track length		Reprojection error (px)		Reconstructionn feature	
Dataset	1	2	1	2	1	2	1	2	1	2
ORB	77.90	73.10	3433	3935	3.64	3.82	1.11	0.86	49196	118185
HAHOG	93.80	94.40	4980	5025	4.33	5.02	0.63	0.59	92825	263324
AKAZE	82.70	77.80	702	1074	3.97	4.53	0.79	0.81	25946	48852
SIFT	92.70	90.70	5766	5408	4.09	5.54	0.39	0.37	108091	287904
NCCFT	94.20	94.40	7238	6702	4.30	5.09	0.38	0.35	124836	284522
PROJ-NCCFT	98.80	98.30	6942	6108	4.68	5.46	**0.28**	**0.27**	123922	288874
Colmap	78.40	77.80	6143	5986	3.62	5.72	0.40	0.39	96864	247279
MODM	99.30	99.20	6265	6387	3.68	5.26	0.39	0.38	102058	273461
Pix4D	**100**	99.60	**16208**	**17306**	–	–	0.75	0.63	117276	**292489**
Ours	**100**	**100**	6942	6108	**5.08**	**6.16**	0.30	0.29	**135483**	291935

a) The best result is marked in bold.

Table 4 Running time (s)

	ORB	HAHOG	AKAZE	SIFT	NCCFT	PROJ-NCCFT	Colmap	MODM	Pix4D	Ours
1	110.82	369.92	121.4	465.44	645.28	558.64	1012.71	1064.28	2440.79	688.45
2	280.25	745.95	219.87	943.36	1348.07	1206.15	2462.43	2255.14	5040.62	1272.78

involved in the reconstruction to 100%. The decrease in feature extraction accuracy for challenging matching images is accompanied by a slight reduction in reprojection error, but still within the acceptable range. Compared to Colmap, the proposed method improves the trajectory length and reconstruction accuracy, which is attributed to the improved completeness of the global optimization. Although the commercial software Pix4D achieves a 100% image reconstruction rate, the proposed method produces results with a significantly lower reprojection error.

In addition, Figure 6 shows the normalized vegetation index (NDVI) rendering of the multispectral image reconstruction results as well as other vegetation indices of local vegetation including ratio vegetation index (RVI) and normalized difference red edge index (NDRE). This demonstrates the value of the application of the results obtained by the asymmetric multispectral stereo imaging system. Multispectral reconstruction results in more bands represent more spectral information allowing exploitation at the 3D level.

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + \mathrm{Red})}, \tag{17}$$

$$\mathrm{NDRE} = \frac{(\mathrm{NIR} - \mathrm{RE})}{(\mathrm{NIR} + \mathrm{RE})}, \tag{18}$$

$$\mathrm{RVI} = \frac{\mathrm{NIR}}{\mathrm{Red}}, \tag{19}$$

where NIR represents near infrared band image, Red represents red band image and RE represents red edge band image.

Figure 7 compares the reconstruction results of the nadir view imaging system with the dual-angle stereo imaging system. As expected, the dual-angle stereo imaging system provides more complete structural and spectral information about the side of the building. Compared with the results of MODM with separate reconstruction of band-by-band images, the proposed method obviously eliminates the ghosting phenomenon in the results and acquires more accurate spectral information. The proposed method achieves satisfactory reconstruction results in terms of visual effects and quantitative evaluation indexes.

4 Conclusion

This paper introduces a dual-angle asymmetric multispectral stereo imaging system and a 3D reconstruction method for the system on UAVs. Compared with existing multispectral stereo imaging systems, this system can acquire more spectral and spatial information, but spectral intensity differences between bands and angular-induced geometric distortions pose greater challenges to the reconstruction method.

The proposed reconstruction method utilizes an adaptive feature extraction method to extract features uniformly across the images of different bands, and a combination of perspective transformation and block re-matching is employed to improve feature matching accuracy and efficiency. Finally, leveraging the multi-view geometry technique, the proposed method successfully acquires MSPC data with multiple band information.

By analyzing the experimental results, the existing methods are insufficient for high quality reconstruction for this multispectral stereo imaging system. The proposed method solves the challenge of reconstructing a dual-angle multispectral system and successfully reconstructs MSPC with more spectral information, which provides data support for applications based on 3D spectral information.

Acknowledgements This work was supported by National Science Fund for Outstanding Young Scholars (Grant No. 62025107) and Open Fund Project of KuiYuan Laboratory (Grant No. KY202423).

References

1 Deng L, Mao Z, Li X, et al. UAV-based multispectral remote sensing for precision agriculture: a comparison between different cameras. ISPRS J Photogramm Remote Sens, 2018, 146: 124–136
2 Qin Z, Li X, Gu Y. An illumination estimation and compensation method for radiometric correction of UAV multispectral images. IEEE Trans Geosci Remote Sens, 2022, 60: 1–12
3 Furukawa F, Laneng L A, Ando H, et al. Comparison of RGB and multispectral unmanned aerial vehicle for monitoring vegetation coverage changes on a landslide area. Drones, 2018, 5: 97
4 Li S, Dian R, Liu H. Learning the external and internal priors for multispectral and hyperspectral image fusion. Sci China Inf Sci, 2023, 66: 140303
5 Sun X, Tian Y, Lu W, et al. From single- to multi-modal remote sensing imagery interpretation: a survey and taxonomy. Sci China Inf Sci, 2023, 66: 140301
6 Brauers J, Schulte N, Aach T. Multispectral filter-wheel cameras: geometric distortion model and compensation algorithms. IEEE Trans Image Process, 2008, 17: 2368–2380
7 Jhan J P, Rau J Y, Haala N. Robust and adaptive band-to-band image transform of UAS miniature multi-lens multispectral camera. ISPRS J Photogramm Remote Sens, 2018, 137: 47–60
8 Brauers J, Aach T. Geometric calibration of lens and filter distortions for multispectral filter-wheel cameras. IEEE Trans Image Process, 2011, 20: 496–505
9 Gallego A, Pertusa A, Gil P, et al. Detection of bodies in maritime rescue operations using unmanned aerial vehicles with multispectral cameras. J Field Robot, 2019, 36: 782–796
10 Pádua L, Adao T, Guimaraes N, et al, Post-fire forestry recovery monitoring using high-resolution multispectral imagery from unmanned aerial vehicles. In: Proceedings of International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Prague, 2019. 301–305
11 Stempliuk S, Menotti D. Agriculture multispectral UAV image registration using salient features and mutual information. In: Proceedings of IEEE International Symposium on Geoscience and Remote Sensing, 2020. 4108–4111
12 Gu Y F, Jin X D, Xiang R Z, et al. UAV-based integrated multispectral-LiDAR imaging system and data processing. Sci China Tech Sci, 2020, 63: 1293–1301
13 Smith P H. Imager for Mars Pathfinder experiment (IMP): a multispectral stereo imaging system. In: Proceedings of the Society of Photo-Optical Instrumentation Engineers, 1998
14 Kazemzadeh F, Haider S A, Scharfenberger C, et al. Multispectral stereoscopic imaging device: simultaneous multiview imaging from the visible to the near-infrared. IEEE Trans Instrum Meas, 2014, 63: 1871–1873
15 Meinen B U, Robinson D T. Mapping erosion and deposition in an agricultural landscape: optimization of UAV image acquisition schemes for SfM-MVS. Remote Sens Environ, 2020, 239: 111666
16 Zainuddin K, Majid Z, Ariff M F M, et al. 3D modeling for rock art documentation using lightweight multispectral camera. Int Arch Photogramm Remote Sens Spatial Inf Sci, 2019, XLII-2/W9: 787–793
17 Li D, Xu L, Tang X, et al. 3D imaging of greenhouse plants with an inexpensive binocular stereo vision system. Remote Sens, 2017, 9: 508
18 Briechle S, Molitor N, Krzystek P, et al. Detection of radioactive waste sites in the Chornobyl exclusion zone using UAV-based lidar data and multispectral imagery. ISPRS J Photogramm Remote Sens, 2020, 167: 345–362
19 Liu B, Chen X G, Guo L P, et al. A MPB-based remote sensing image 3D reconstruction system. Optik-Int J Light Electron Opt, 2015, 126: 1994–1998
20 Zhao L, Wang H, Zhu Y, et al. A review of 3D reconstruction from high-resolution urban satellite images. Int J Remote Sens, 2023, 44: 713–748
21 Zhang Z, Zhang L, Tong X, et al. A multilevel point-cluster-based discriminative feature for ALS point cloud classification. IEEE Trans Geosci Remote Sens, 2016, 54: 3309–3321
22 Hosoi F, Umeyama S, Kuo K. Estimating 3D chlorophyll content distribution of trees using an image fusion method between 2D camera and 3D portable scanning lidar. Remote Sens, 2019, 11: 2134
23 Jurado J M, López A, Pádua L, et al. Remote sensing image fusion on 3D scenarios: a review of applications for agriculture and forestry. Int J Appl Earth Obs GeoInf, 2022, 112: 102856
24 Remondino F, El-Hakim S. Image-based 3D modelling: a review. Photogramm Record, 2006, 21: 269–291
25 Sattler T, Leibe B, Kobbelt L. Efficient & effective prioritized matching for large-scale image-based localization. IEEE Trans Pattern Anal Mach Intell, 2017, 39: 1744–1756
26 Tafti A P, Baghaie A, Kirkpatrick A B, et al. A comparative study on the application of SIFT, SURF, BRIEF and ORB for 3D surface reconstruction of electron microscopy images. Comput Methods BioMech BioMed Eng-Imag Vis, 2018, 6: 17–30
27 Fan B, Kong Q, Wang X, et al. A performance evaluation of local features for image-based 3D reconstruction. IEEE Trans Image Process, 2019, 28: 4774–4789
28 Xie X, Yang T, Li D, et al. Hierarchical clustering-aligning framework based fast large-scale 3D reconstruction using aerial imagery. Remote Sens, 2019, 11: 315

29 Seifert E, Seifert S, Vogt H, et al. Influence of drone altitude, image overlap, and optical sensor resolution on multi-view reconstruction of forest images. Remote Sens, 2019, 11: 1252

30 Schonberger J L, Frahm J M. Structure-from-motion revisited. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, 2016. 4104–4113

31 Campo F B, Ruiz F L, Sappa A D. Multimodal stereo vision system: 3D data extraction and algorithm evaluation. IEEE J Sel Top Signal Process, 2012, 6: 437–446

32 Wang L, Xiong Z, Shi G, et al. Simultaneous depth and spectral imaging with a cross-modal stereo system. IEEE Trans Circ Syst Video Technol, 2018, 28: 812–817

33 Vong A, Matos-Carvalho J P, Toffanin P, et al. How to build a 2D and 3D aerial multispectral map? — All steps deeply explained. Remote Sens, 2021, 13: 3227

34 Kwan C, Chou B, Ayhan B. Enhancing stereo image formation and depth map estimation for mastcam images. In: Proceedings of IEEE Ubiquitous Computing, Electronics and Mobile Communication Conference, 2018. 566–572

35 Wang C, Gu Y, Li X. A robust multispectral point cloud generation method based on 3-D reconstruction from multispectral images. IEEE Trans Geosci Remote Sens, 2023, 61: 1–12

36 Zhang Z, Zhu L. A review on unmanned aerial vehicle remote sensing: platforms, sensors, data processing methods, and applications. Drones, 2023, 7: 398

37 Snavely N, Seitz S M, Szeliski R. Photo tourism: exploring photo collections in 3D. ACM Trans Graph, 2006, 25: 835–846

38 Schonberger J L, Zheng E L, Frahm J M, et al, Pixelwise view selection for unstructured multi-view stereo. In: Proceedings of European Conference on Computer Vision, 2016. 501–518

39 Feng R, Shen H, Bai J, et al. Advances and opportunities in remote sensing image geometric registration: a systematic review of state-of-the-art approaches and future research directions. IEEE Geosci Remote Sens Mag, 2021, 9: 120–142

40 Bansal M, Kumar M, Kumar M. 2D object recognition: a comparative analysis of SIFT, SURF and ORB feature descriptors. Multimed Tools Appl, 2021, 80: 18839–18857

41 Ma J L, Chan J C W, Canters F. Fully automatic subpixel image registration of multiangle CHRIS/Proba data. IEEE Trans Geosci Remote Sens, 2010, 48: 2829–2839

42 Wu Y, Ma W, Su Q, et al. Remote sensing image registration based on local structural information and global constraint. J Appl Rem Sens, 2019, 13: 1

43 Kern J P, Pattichis M S. Robust multispectral image registration using mutual-information models. IEEE Trans Geosci Remote Sens, 2007, 45: 1494–1505

44 Cole-Rhodes A A, Johnson K L, LeMoigne J, et al. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. IEEE Trans Image Process, 2003, 12: 1495–1511