RESEARCH ON CLUSTERING MATCHING STRATEGY OF 3D RECONSTRUCTION BASED ON IMAGE FUSION FEATURES

Tiebo SUN [1,2], Jinhao LIU [1,*], Jiangming KAN [1], Tingting SUI [1], Jiang LI [1], Miaoxin JI [1]

In an attempt to remedy the low efficiency of many 3D reconstructions of high-resolution sequence images from UAV tilt photography for forestry surveys, based on the existing motion recovery structure algorithm, a reconstruction algorithm using UAV image fusion features is proposed in this paper. First, the GPS/IMU information of the UAV image was extracted. Second, the UAV image was downsampled and blocked, and the color invariant standard deviation of each image block was calculated simultaneously. Then, the texture features, such as the angular second moments of the downsampled images, were calculated based on GLCM theory by using the above feature components. The L^2 norm was normalized, and then the image description vector was constructed. In addition, the p-stable distribution LSH algorithm was used to cluster the matched images and the clustering results were used to perform image matching. The method described in this paper can effectively reduce the image matching range in the reconstruction process and improve reconstruction efficiency. The experimental results showed that the algorithm mentioned in this paper improved the reconstruction efficiency by approximately three times, and the number of dense point clouds was only 3% lower than that of the global match reconstruction algorithms. The target details, such as the trees in the core area of the slope photographs, were well restored. The accuracy of the point cloud data obtained by this algorithm was more than 85% in comparison with the measured data, with a crown width R^2 of 0.972, and a tree height R^2 of 0.979. The method provided in this paper can meet the 3D reconstruction requirements of forestry surveys and has the ability to guaranteeing accuracy.

Keywords: forestry survey, unmanned aerial vehicle, three-dimensional reconstruction, fusion feature, image matching, image clustering

1. Introduction

A vital index reflecting the main forest characteristics is the biomass of forest ground vegetation. There is a positive correlation between plant biomass and height [1]. Image-based 3D reconstruction technology has advanced rapidly

[1] School of Technology, Beijing Forestry University, Beijing 100083, China
[2] School of Mechanical and Electrical Engineering, Jiangsu Food & Pharmaceutical Science College, Huai'an 223003, China
***** Correspondence author: Jinhao Liu, email: liujinhao@vip.163.com

due to the continuous progress and improvement of feature point detection, matching algorithms [2,3], self-calibration algorithms [4-6], structure from motion [7,8], and multiview stereopsis [9,10]. It has been widely used in forestry resource investigations, stand studies, and 3D reconstruction of standing trees. UAVs have the advantages of flexible views and quick sampling because they fly at low altitudes. A large number of uncalibrated image sequences with a continuous view of the same scene can be obtained by using a UAV. At the same time, the three-dimensional information of the scene can be obtained by combining it with SFM technology. A new solution to increase the speed of 3D reconstruction of UAV sequence images is urgently needed because of the high resolution and overlap of UAV images and the low efficiency of SFM processing.

The most time-consuming step of the SFM processing of UAV sequence images is image feature matching and bundle adjustment. To recover more details during SFM 3D reconstruction, an exhaustive strategy is adopted and global matching occurs during the calculation of the constrained relationships of the sequence images. The time complexity can be obtained as O (n^2), where O is the number of images. Shum H and other analyses indicated that the time complexity in the process of bundle adjustment reached O (n^4). The effective matching relationship of the image and the removal of redundant matching calculations are the primary objectives of efficient 3D reconstruction of UAV sequence images.

Snavely and other authors [11] have used a skeleton extraction strategy based on a feature to reconstruct groups of images to improve processing efficiency, which effectively improves the efficiency of reconstruction; however, the picture features were not fully considered. Jared Heinly and other authors [12] proposed a reconstruction computing framework for SFM based on streams. The algorithm was suitable for large-scale 3D reconstruction of internet images, but it failed to also take image features into account. Farenzena M and other authors [13] created an image of a tree by measuring the overlapping distances and following the order of leaves to roots during reconstruction. This method decomposed the whole reconstruction problem into smaller instances, solved them separately and combined them. The method has lower computational complexity, but there was redundancy in the distance matching of small scenes. AliAkbarpour and the other authors [14] used the shooting time constraints to reduce the image matching complexity and used the continuous frame strategy to match image features, which significantly improved the matching efficiency. KunSun and other authors [15] proposed centrally driven image set partitioning and clustered images by calculating the overlap degree of images and introduced image relationships into a three-dimensional reconstruction, which improved the efficiency of the three-dimensional reconstruction but did not consider other features of the images. Xu Zhihua and other authors [16] proposed a reconstruction method for the topological relationships of images based on GPS information, limited the image

matching to topological relationships of the images during the process of 3D reconstruction, which reduced the matching time; the situation of the same camera position but a completely different shooting angle was not considered. Song Zhengxi and other authors [17] divided images into blocks and constructed an image kd tree by extracting the SIFT features and then matched the images. However, the initial image clustering still uses SIFT features, and the clustering efficiency was not very high.

To solve the above problems, this paper proposed a reconstruction matching strategy that combines POS information and the color texture information of UAV images. The pre-reconstruction images were clustered according to the characteristics of the UAV image to avoid the exhaustive matching process of global match reconstruction algorithms. The results showed that the proposed method can significantly improve the efficiency of three-dimensional reconstruction of oblique photographic images, and the reconstruction details of trees and other objects had better accuracy.

2. Clustering Algorithms based on UAV Sequence Images

The sequence images taken by the UAV had position information that was accurate to 10 meters and attitude information that was accurate to 5 degrees. At the same time, its course overlap and side overlap were relatively high. Because of the above characteristics, the image clustering in this study considered both the pose/attitude information and its characteristics, and the steps were mainly divided into GPS/IMU feature extraction, color and texture feature extraction, eigenvector generation and clustering.

2.1. Image features

(1) Position and attitude characteristic components. The UAV's airborne GPS/IMU system recorded the UAV's spatial information and attitude information when images were formed at the time of exposure and written into the EXIF file. The output pose information from the airborne GPS/IMU system was as follows: $C_i = [R_i | t_i]$, where R_i is a 3-order rotation matrix, and t_i is a three-dimensional space vector containing the position and attitude information of the camera. The position and attitude information obtained by the UAV flight control system includes the yaw angle, roll angle and pitch angle (Φ, Θ, Ψ) of the IMU body coordinate system relative to the navigation coordinate system. The outer azimuth element (ψ, ω, κ) of the image space coordinate system in the ground measurement coordinate system can be determined by a coordinate transformation. The image position and attitude vector components were composed of the converted IMU information, and the UAV position coordinate

vectors in the coordinate system of earth-centered earth-fixed can be expressed as $G_r = [x_r, y_r, z_r]$.

(2) Color feature components. Color is the global image feature that can provide the most intuitive information during the process of image classification. This study used color information as an image feature component because of the abundant color information provided by unmanned aerial vehicle (UAV) images. There are many expressions of color. We calculated the color invariants along with the features of illumination and rotation invariants based on the Kubelka-Munk theory and used the color invariants to measure the color similarity component.

According to the Kubelka-Munk theory, the model of spectral radiation characteristics of objects in the opaque state is as follows [18]:

$$E(\lambda, \tilde{x}) = e(\lambda, \tilde{x})[1 - \rho_f(\tilde{x})]^2 R_\infty(\lambda, \tilde{x}) + e(\lambda, \tilde{x})\rho_f(\tilde{x}) \tag{1}$$

where λ is the wavelength of the light, $\tilde{x}$ is the position of the imaging plane, $e(\lambda, \tilde{x})$ is the intensity of the spectrum, $\rho_f(\tilde{x})$ is Fresnel's reflection coefficient under $\tilde{x}$, $R_\infty(\lambda, \tilde{x})$ is the reflectivity of the materials, and $E(\lambda, \tilde{x})$ is the reflection spectra at the observation sites. Normally, $e(\lambda, \tilde{x})$ is relevant to the position only. $e(\lambda, \tilde{x})$ was rewritten as $i(\tilde{x})$, which changed (1) into the following:

$$E(\lambda, \tilde{x}) = i(\tilde{x})\{[1 - \rho_f(\tilde{x})]^2 R_\infty(\lambda, \tilde{x}) + \rho_f(\tilde{x})\} \tag{2}$$

The color invariant K can be obtained by dividing the first derivative and the second derivative of λ from the above formula as follows:

$$K = \frac{E_\lambda}{E_{\lambda\lambda}} = \frac{\partial E / \partial \lambda}{\partial^2 E / \partial^2 \lambda} = \frac{\partial R_\infty(\lambda, \tilde{x}) / \partial \lambda}{\partial^2 R_\infty(\lambda, \tilde{x}) / \partial \lambda^2} = f[R_\infty(\lambda, \tilde{x})] \tag{3}$$

The *K* value was only related to the object reflectivity from formula (3). We used the linear transformation and Gauss color model to obtain the conversion relationship of $(E, E_\lambda, E_{\lambda\lambda})$ and the RGB components, following the CIE-1964-XYZ standard and human visual characteristics as follows:

$$\begin{pmatrix} E \\ E_\lambda \\ E_{\lambda\lambda} \end{pmatrix} = \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.60 & 0.17 \end{pmatrix} \times \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{4}$$

The steps of constructing a color descriptor were as follows: first, the image was segmented into 4×4 areas according to the size of the image, after which the color invariant K standard deviation of the feature points distributed in 4×4 regions was counted, and finally, a 16-dimensional color descriptor was formed as follows: $C = [C_1, C_2, ...C_i, ..., C_{16}]$. C_q was constructed in the corresponding region *q* with $m \times n$ pixels, and C_q can be defined as follows:

$$C_q = \sqrt{\frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (K(i, j) - \mu)^2} \tag{5}$$

where $K(i, j)$ is the color invariants of line *i* and column j in region *q* and μ is the mean value of color invariants in region *q*.

(3) Texture feature description vectors. Texture is essential information for describing image features. The gray-level cooccurrence matrix (GLCM) is a very effective method for describing image texture features, as shown by relevant research. The GLCM is a second-order statistical feature that can describe image texture, and it represents image texture through joint probability density functions between different locations. It can reflect the overall distribution characteristics of the same brightness and the location distribution characteristics of areas with similar brightness values simultaneously. If a two-dimensional digital image with the gray level N_g and size $M \times N$ is expressed by $f(x, y)$, then the image GLCM can be represented as follows:

$$G(i, j) = \frac{\mathrm{F}(i, j)}{\sum_{i=0}^{N_g} \sum_{j=0}^{N_g} \mathrm{F}(i, j)} \tag{6}$$

where $F(i, j)$ is the points logarithm of $f(x, y)$, which satisfies the $\{(x_1, y_1), (x_2, y_2) \in M \times N | f(x_1, y_1) = i, f(x_2, y_2) = j\}$ condition in image $f(x, y)$. Further, *G* is the N_g-order square matrix. The distance between two points (x_1, y_1) and (x_2, y_2) is *d*, and the angle of the line between the two points and the coordinate axis is θ. The components of the GLCM can be calculated from *d* and θ as follows: $G(i, j, d, \theta)$.

Various features describing image texture can be obtained according to the GLCM components, and fourteen image texture features can be defined based on the GLCM. In this paper, eight texture features and the texture similarity between images were selected for calculation, including the correlation (*COR*), entropy (*ENT*), moment of inertia (*CON*), heterogeneity (*DIS*), angular second moment (*ASM*), inverse difference moment (*HOM*), average value (*u*) and standard deviation (σ).

The image collected by the UAV was first sampled eight times when calculating the GLCM because of the large number of calculations. $d = 1, \theta \{0^o, 45^o, 90^o, 135^o\}$ and the GLCM in four directions was calculated, and then 32-dimensional vectors that described the image texture features were obtained by calculating eight texture feature values in four directions.

(4) Fusion feature description vector. To eliminate the dimension influence between features and facilitate the comparison and matching work of the fusion

feature descriptors in the later stage of fusion work, it was necessary to standardize the eigenvalues of the three described subcomponents because of the different dimensions in the GPS/IMU feature, color feature and texture feature. In this paper, the *L2* norm method was used to standardize the three sets of feature data above. The *L2* norm of the vector $X(X_1, X_2, \ldots, X_n)$ was defined as follows:

$$norm(x) = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} \tag{7}$$

where 1 was chosen as the upper limit to standardize the data and the mapping x' of x can be established according to the following formula:

$$1 = norm\left(x'\right) = \frac{\sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}}{norm(x)} \tag{8}$$

The following conclusion can be drawn:

$$x_i' = \frac{x_i}{norm(x)} \tag{9}$$

2.2. p-Stable Locally Sensitive Image Hash Clustering Algorithm

The high-dimensional vector had good performance in the locality-sensitive hashing (LSH) [19]. Different hash effects can be produced by different hash function clusters according to the characteristics of LSH [20].

If under the condition $p \geq 0$, n random variables $X_1, X_2, \ldots, X_n$ in D are compared to n real numbers $a_1, a_2, \ldots, a_n$, then the following is true:

$$P(\sum_{i=1}^{n} a_i X_i) = P(\sqrt[p]{\sum_{i=1}^{n} |a_i|^p}\, X) \tag{10}$$

Then, D can be defined as a p-stable distribution.

P-stable distributions can calculate the Euclidean distance between vectors directly and reduce the high-dimensional features dimension at the same time according to their natures. Specifically, the random function of the p-stable distribution was used to generate an independent random vector a. For the 54-dimensional image fusion feature vector v, $a \cdot v$ can be used to estimate $||v||^p$.

The function cluster of LSH under the p-stable distribution was defined as follows:

$$h_{a,b}(v) = \left\lfloor \frac{a \cdot v + b}{w} \right\rfloor \tag{11}$$

where a is a d-dimensional vector generated by a random function that satisfies the requirements of a p-stable distribution. Letter b is a random real number with a range of $[0, w]$. Specifically, the fused eigenvector v was mapped to vector a; at

the same time, *b* was used to correct the deviation caused by the mapping process, then quantitative disposal was conducted with interval *w*. In this way, similar vectors can be mapped to the same intervals.

$c = \| v_1 - v_2 \|_p$ $f_p(t)$ was set as the absolute value of the p-stable distribution probability density function of two arbitrary vectors v_1 and v_2 in the dataset, so $|a \cdot v_1 - a \cdot v_2| < w$ was the mapping of eigenvectors v_1 and v_2 to *a*, that is, $|(v_1 - v_2) \cdot a| < w$; further, the following was obtained:

$$\| v_1 - v_2 \|_p \; X = | cX | < w \tag{12}$$

The collision probability between v_1 and v_2 obtained by formula (11) was as follows:

$$p_r(c) = p_r(h(v_1) = h(v_2)) = \int_0^w \frac{1}{c} f_p\left(\frac{t}{c}\right)\left(1 - \frac{t}{w}\right) dt \tag{13}$$

From formula (13), the probability decrease with the increase in *c* in the case of *w* was unchanged. A set of dimension-reduced hash values can be obtained by sequence image fusion features through the LSH algorithm p-stable distribution, which enters $h_1, h_2, \ldots, h_k$ into the hash table.

2.3. Algorithmic Flow

The motion recovery structural algorithm based on the fused features in this paper includes four steps: feature generation, image clustering, feature matching, and motion recovery structure. The structure of motion recovery was the same as that of the SFM algorithm. The flow chart of the fusion feature matching algorithm for the UAV sequence images was as follows:

Step 1. The GPS/IMU values were extracted from the aerial sequence images, and the six-dimensional positions and pose feature components were obtained.

Step 2. The 16-dimensional color feature component of the color invariant component was extracted from the image in 4×4 blocks through 8-fold down sampling and 4-dimensional 32-dimensional texture feature descriptor extraction. The 54-dimensional fusion feature vectors of UAV sequence images were obtained by standardized data processing and combined.

Step 3. To construct the *L* functions, an independently randomly selected hash function in the LSH function cluster of the p-stable distribution from Formula (11) was used.

Step 4. The 54-dimensional feature vectors of the UAV image sequence were mapped by the LSH function. The high-similarity images were mapped into a hash bucket.

Step 5. The Euclidean distances of the vectors in the hash bucket were calculated one by one. The distance calculation results were arranged, and the first *k* images were taken to form the set of images to be matched.

Step 6. The matching relationship between the query image and the image set to be matched was calculated based on the hash value. The matching relationships were recorded and the process of the reconstruction matching operation was substituted.

According to the algorithm, there were *k* images matching each image, but the number of images to be matched was less than *k*. In this case, it was considered that all images matched the current image.

3. Experimental Results and Discussion

The experimental image data of the two scenes, the sole tree species nursery area and the complex woodland with water in the core area were obtained (the SFM algorithm is more likely to lose details in the process of water surface reconstruction), and the DJI Mavic Air UAV was used to test and verify the effectiveness of the algorithm proposed in this paper. Experimental site 1 included farmland, low trees, and roads. The altitude was set at 25 meters, and images within the core area of 50×50 m^2 of low trees were taken in five directions. The image size was 4056 *pixels*×3040 *pixels*, the pixel size was 1.5 μm, and the total number of scene images was 870. Experimental site 2 includes landscape trees, shrubs and water surfaces. The altitude was set at 80 meters, and images were taken in five directions within the core 500×500 m^2 area of low trees. The total number of images for the site was 582. Experimental site 3 was from an open dataset (rising valley NIR) provided by OpenDroneMap and was selected to test the algorithm. The UAV had a flight height of *430 m*, including trees, low vegetation and buildings. There were 566 images in the dataset, and each image was 4048 *pixels* × 3048 *pixels*. The experimental environment was as follows: Linux 16.04 64-bit operating system with an Intel i9-9900K Processor and 64G of memory. The degree of frontlap and sidelap of the two flights was 80%.

The image matching module of the SFM reconstruction framework was replaced with the algorithm presented in this paper. The CMVS/PMVS module was used for dense point cloud reconstruction, and the two groups of experimental data above were used for three-dimensional reconstruction. This paper compared the global match SFM algorithms Bundler [21] and VSFM [22] and analyzed the reconstruction efficiency and reconstruction effect. Table 1 shows the POS information of some of the experimental data.

Table 1

Partial POS information in experimental data

PhotoID	X	Y	Z	Pitch angle	Rolling angle	Navigation angle
0072	33.44956207	119.02542877	80.5	45.09776	0.50770	0.50595
0073	33.44957352	119.02534485	80.5	45.09924	0.29517	0.29414
0074	33.44958496	119.02520752	80.6	45.09956	0.22428	0.22350
0075	33.44960403	119.02503204	80.6	44.99924	0.29497	0.29454
0076	33.44962692	119.02483368	80.5	45.09924	0.29489	0.29394

3.1. Matching Times

The image matching strategy is exhaustive under the global match SFM framework. Specifically, $C_n^2 = n \times (n\text{-}1)/2$. The image of this algorithm is matched with a fixed number of datasets. For $N \times n$, the size of the combined test area and the quantity of images in this experiment, set *k* was 30 in site 1, and *k* was 25 in site 2. The matching times of the three sites are shown in Fig. 1.

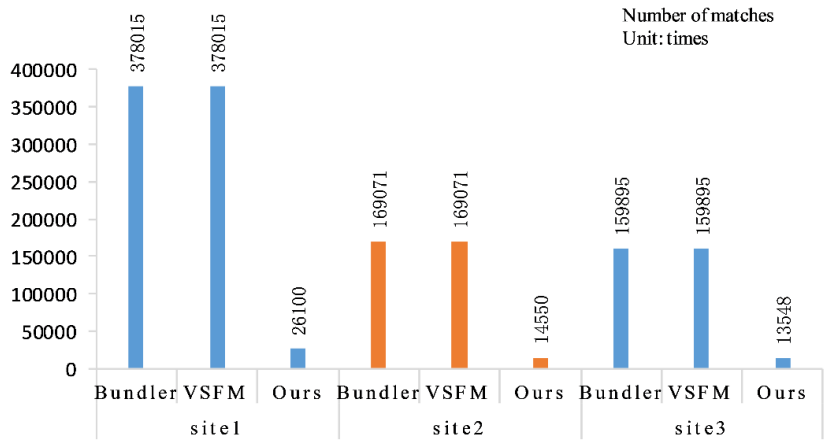


Fig. 1. Matching times of the three sites

3.2. Time of Reconstruction

The reconstruction times of this method and other methods in different stages are shown in Table 2. The image of site 1 contains more details because of the lower flight altitude and because more pictures were taken. The feature extraction, matching, and reconstruction were more time-consuming than those of site 2. The time wastage was basically the same in the link of feature extraction because the SIFT feature extraction method was adopted in all cases. The time consumption was high during feature matching after the feature points were extracted because the exhaustive method was used with the traditional global

match SFM algorithm. The efficiency of this algorithm in feature matching was much higher than that of the global match algorithm. The total time consumed was proportional to the number of images. The matching times for site 1 and site 2 were more than 5 times less than that of the global match algorithm. The efficiency of three-dimensional reconstruction was increased by approximately three times. At the same time, the implementation of the algorithm without hash mapping was verified. The image was directly subjected to Euclidean distance clustering through the fusion feature vector, and the *k* value was unchanged. It can be seen that when there was no hash map to index, the feature matching time increased considerably because the Euclidean distance between two images had to be calculated.

Table 2

Time of different stages of reconstruction using ours method and other methods

(Unit: *min*)

Algorithm time consuming	site1			site2			site3		
	Bundler	VSFM	Ours	Bundler	VSFM	Ours	Bundler	VSFM	Ours
Feature extraction	45.6	40.5	40.6	30.4	30.1	30.4	32.4	32.2	32.2
Feature matching	170.8	158.4	24.3	114.2	98.6	16.8	126.6	121.5	17.6
Non-LSH	-	-	70.5	-	-	50.6	-	-	52.3
Reconstruction	30.3	28.6	22.5	17.7	14.7	12.6	19.2	15.9	14.6
Total time	246.7	227.5	87.4	162.3	143.4	59.8	178.2	169.6	64.4

Five hundred images were matching time-consuming curves separately in site 1. Based on the algorithms in this paper and testing a single image with the global match algorithms, the curve is shown in Fig. 2. The total matching time of a single image in the global match algorithms increased with the number of images, as shown in Fig. 2. This is because in the exhaustive method, a single image needs to be matched with all other images, and the different number of features in the image to be matched increases the matching time, leading to a larger change in the curve. The matching time of this algorithm does not increase with the number of images because the number of matches remains unchanged. However, the matching time changes slightly due to the different numbers of image databases and image features.

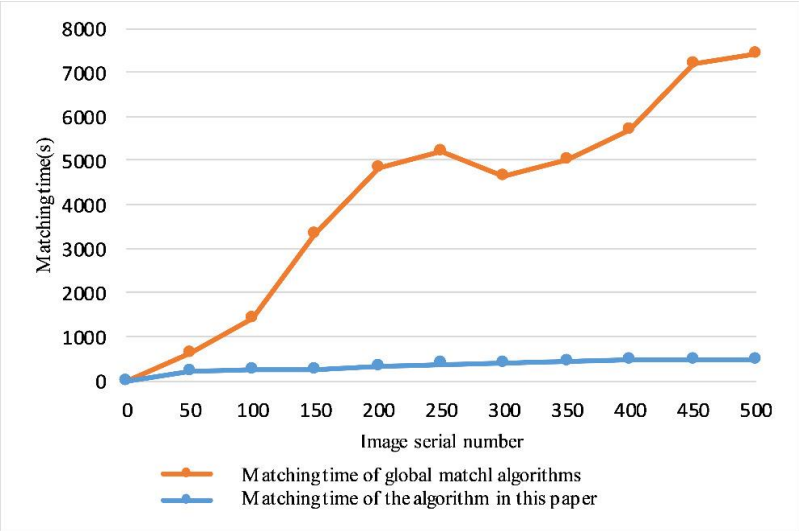


Fig. 2. Matching time of single image

The 102 image of site 3 and its 24 matching pairs were output as shown in Fig. 3.

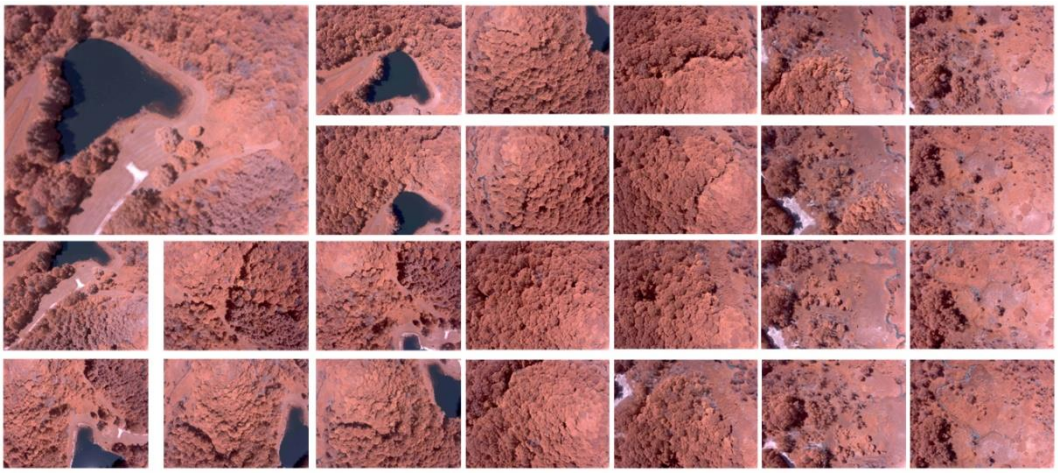


Fig. 3. Pair of matched images

3.3. Effect of the Dense Point Cloud Reconstruction

On the basis of the reconstruction completed in the last step and the obtained sparse point clouds and the dense point clouds from the two scenarios generated by CMVS/PMVS, the number of point clouds obtained is shown in Fig. 4. The proposed algorithm was slightly reduced compared with those of the Bundler and VSFM judging from the number of point clouds. Site 1 was reduced

by approximately 2.5%, site 2 was reduced by approximately 2.8%, and site 3 was reduced by approximately 2.3%.

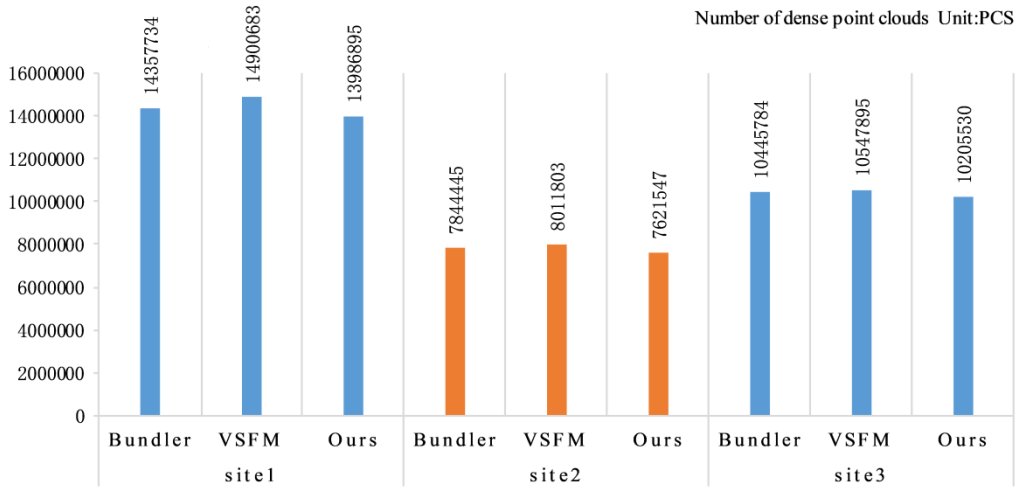
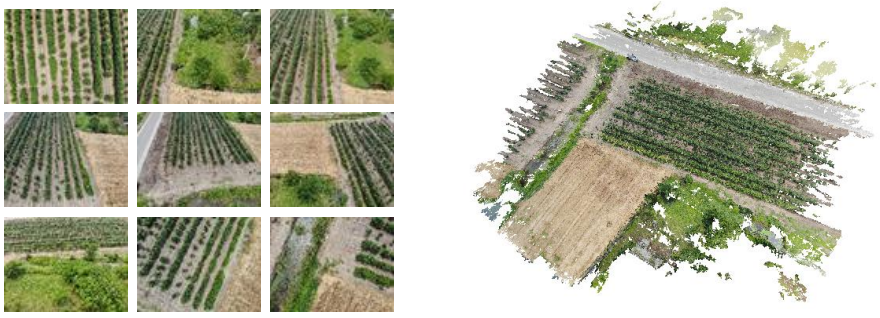


Fig. 4. Number of dense point clouds in three sites

All three algorithms completed the 3D reconstruction of the target area by combining the method in this paper with the other two global matching SFM methods from the final result of the dense point clouds. The reconstructions are shown in Figs. 5-7. Site 1 contained water bodies; as a result, the three algorithms did not effectively recreate the water, which is also a common problem with the SFM algorithm. The point cloud details in the oblique photographic boundary region were not as good as those of the two global match algorithms compared with the algorithm proposed in this paper.



(a) (b)

Fig. 5. Site 1 Dense Point Cloud. (a) A typical picture of Site 1; (b) Site 1 dense point cloud under Bundler

(c) (d)

Fig. 5. Site 1 Dense Point Cloud. (c) Site 1 dense point cloud under VSFM; (d) Site 1 dense point cloud under Ours method

(a) (b)

(c) (d)

Fig. 6. Site 2 Dense Point Cloud. (a) Site 2 typical pictures; (b) Site 2 dense point cloud in Bundler; (c) Site 2 dense point cloud in VSFM; (d) Site 2 dense point clouds under Ours method

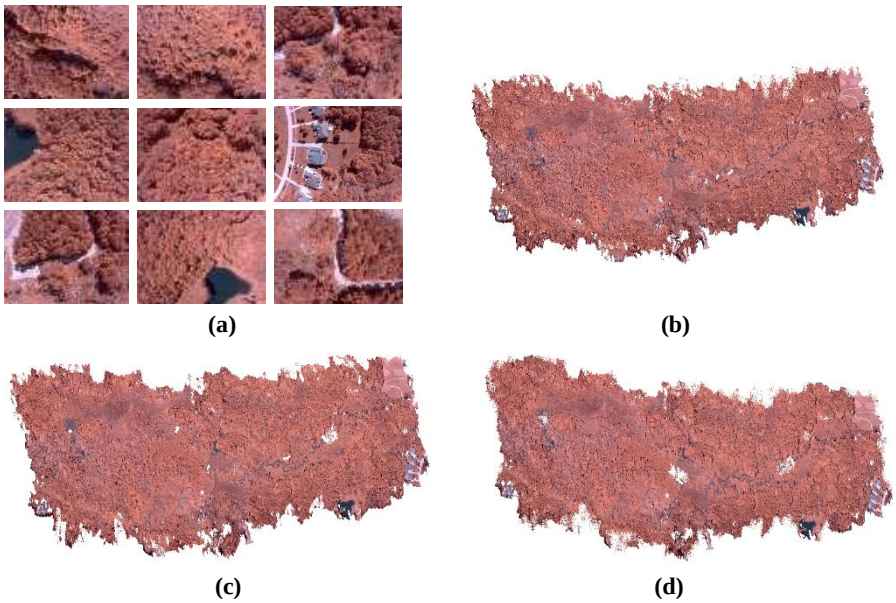


Fig. 7. Site 3 Dense Point Cloud. (a) A typical picture of Site3; (b) Site3 dense point cloud under Bundler; (c) Site 3 dense point cloud under VSFM; (d) Site3 dense point cloud under Ours method

One of the reasons was that the peripheral area imaged through tilt photography had fewer matching features due to the single photographic angle, and another reason was the failure of the partial image association in the boundary region due to the dramatic IMU information changes induced by the sudden turns of the UAV.

The ability of the algorithm to reconstruct the detailed of the core area detail through slope photography was comparable to that of the two global match algorithms.

3.4. Accuracy of the Dense Point Cloud Reconstruction

Ten trees in the center of the tilted photographic area under site 1 and site 2 were selected, and manual measurement data of the tree crown width and height were compared with the point cloud data to verify the accuracy of the reconstructed dense point cloud. The point cloud size was obtained by calculating the minimum bounding rectangle after manual segmentation of the individual trees. The crown size was half of the sum of the length and width of the outer rectangle, and the tree height was the height of the outer rectangle, the point cloud of single tree and external rectangle are shown in Fig. 8.

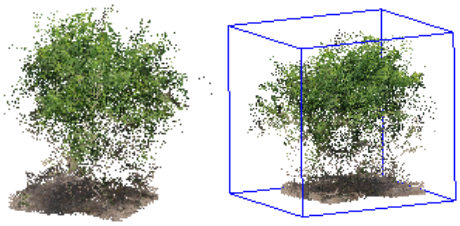


Fig. 8. Segmented individual tree point clouds and circumscribed rectangle diagram of the point cloud

The measured crown amplitude data and point cloud crown amplitude data were obtained as shown in Table 3:

Table 3

Measured crown amplitude data measured and point cloud crown amplitude data

(Uint: *m*)

		Measured data		Bundler point cloud data		VSFM point cloud data		Point cloud data based on ours method	
		Crown width	Tree height	Crown width	Tree height	Crown width	Tree height	Crown width	Tree height
Site 1	1	0.74	1.98	0.66	2.12	0.94	1.72	0.63	2.27
	2	0.62	1.67	0.54	1.45	0.71	1.86	0.72	1.91
	3	0.87	2.09	0.92	1.78	0.74	1.81	0.91	1.78
	4	0.93	2.21	1.02	2.42	0.81	2.46	0.80	1.95
	5	0.58	1.53	0.50	1.45	0.5	1.44	0.50	1.71
	6	0.63	1.87	0.54	1.67	0.55	1.65	0.72	2.14
	7	0.57	1.74	0.64	1.59	0.58	2.00	0.51	1.51
	8	0.86	2.02	0.92	2.3	0.95	2.02	0.74	1.97
	9	0.89	1.96	0.76	1.85	1.01	2.11	0.99	2.25
	10	0.61	1.24	0.52	1.14	0.52	1.09	0.53	1.1
Site 2	11	3.52	8.76	3.21	9.46	3.81	7.81	3.14	9.27
	12	4.69	11.54	4.81	10.78	3.99	12.51	3.99	12.57
	13	4.01	10.12	4.22	9.21	3.62	9.17	4.29	9.15
	14	3.21	9.15	2.91	8.46	3.48	9.43	2.84	9.40
	15	4.12	10.15	4.31	10.47	3.51	9.09	3.53	9.10
	16	4.53	11.25	4.03	10.23	4.79	10.28	4.78	10.19
	17	3.56	8.36	3.14	7.78	3.79	8.66	3.03	8.57
	18	3.74	7.41	3.95	7.88	3.28	7.99	3.29	6.49
	19	3.11	6.66	3.32	6.01	2.82	6.05	2.65	7.06
	20	3.95	9.64	3.59	10.38	3.56	10.15	3.98	8.97

The absolute error of the 3D reconstructed point cloud can be obtained from the measured data and point cloud data, as shown in Fig. 9. The accuracy analysis of the point cloud data is shown in Fig. 10.

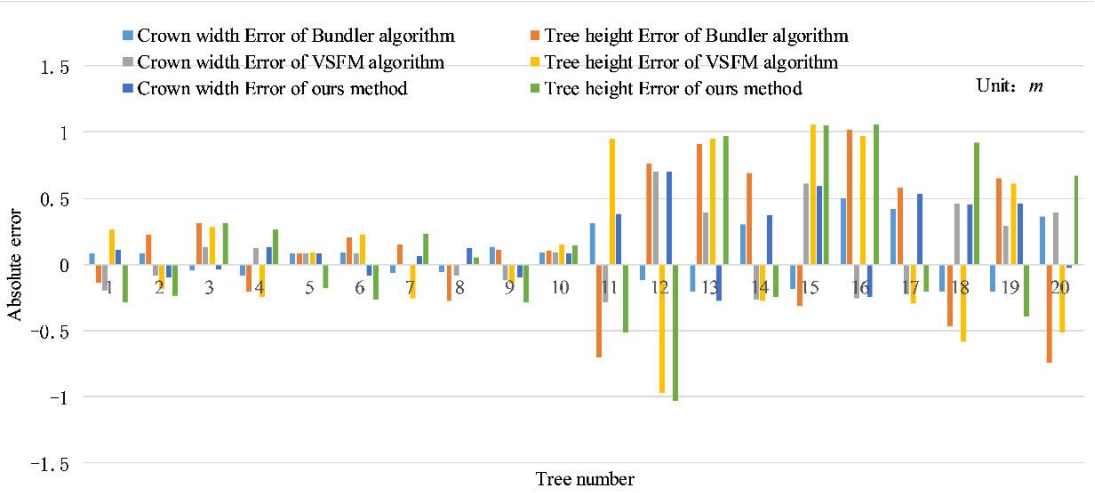
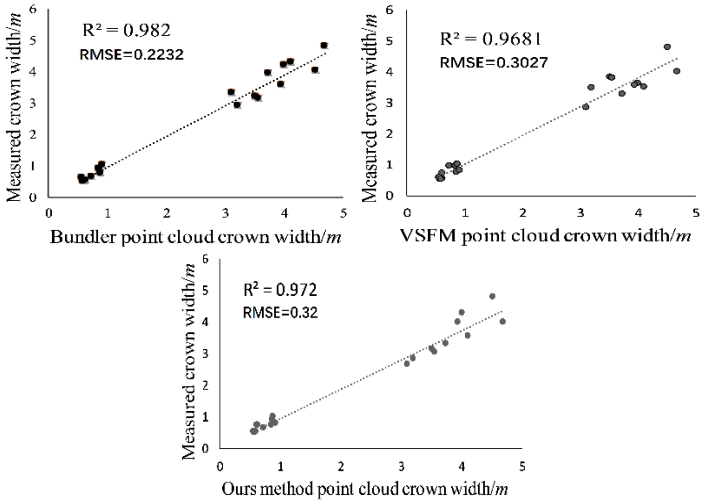
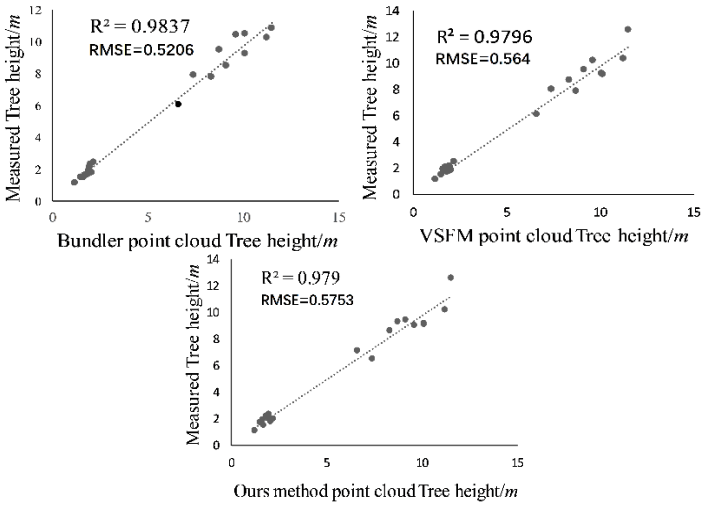


Fig. 9. Absolute Error of the crown width and tree height data from the point cloud



(a) (b) (c)

Fig. 10. Accuracy analysis of the crown width and tree height data from the measured data and the data extracted from the point cloud. (a) Crown width accuracy of Bundler point cloud; (b) Crown width accuracy of VSFM point cloud; (c) Crown width accuracy of Ours point cloud;



(d) (e) (f)

Fig. 10. Accuracy analysis of the crown width and tree height data from the measured data and the data extracted from the point cloud. (d) Tree height accuracy of Bundler point cloud; (e) Tree height accuracy of VSFM point cloud; (f) Tree height accuracy of Ours point cloud

As shown in Fig. 9, the absolute error of the point cloud data obtained by this algorithm was smaller than that of the two global match algorithms, whether from the crown size data or tree height data, the crown width maximum error rate of the Bundler algorithm was 14.75%, with an RMSE of 0.2232, and the tree height maximum error rate was 14.83%, with an RMSE of 0.5206. The crown width maximum error rate of the VSFM algorithm was 14.94%, with an RMSE of 0.3027, and the tree height maximum error rate was 14.94%, with an RMSE of 0.564. The crown width maximum error rate of the algorithm described in this paper was 14.93%, with an RMSE of 0.32, and the tree height maximum error rate was 14.83%, with an RMSE of 0.5753. The error ranges of the three algorithms were consistent, and the results showed that this algorithm can guarantee the accuracy of the point cloud by improving matching efficiency. The RMSE of this algorithm was slightly larger than those of the two global matching algorithms because the details of the point clouds on the tree contours were lost, which was caused by the reduction in the number of matched images. Furthermore, the correlation between the algorithm described in this paper, the two global matching algorithms and the measured data was good, as shown in Figure 10, and the R^2 value was above 0.96. The accuracy of the point cloud data obtained by this algorithm can meet the needs of large-scale forestry investigations, but the data error data of the three algorithms in site 1 under low trees were larger than those in site 2. This also conformed to the technical characteristics of the SFM algorithm. The flight altitude, tree height and canopy density affected the

acquisition of forest point cloud data by the low-altitude tilt photography of the UAVs.

6. Conclusions

Aiming at the low efficiency of so many inclined image data 3D reconstructions in forestry surveys, an image clustering algorithm combined with UAV image features and pose features was proposed in this paper. This algorithm can realize image similarity clustering before matching, utilize GPS/IMU information and image color texture information of UAVs, and avoid exhaustive clustering with global match algorithms. For the forestry survey, two scenes of single tree species nursery areas and complex woodlands with water areas were selected to collect the tilt image data, and the effect, matching time and number of dense point clouds were compared between the two scenarios. The efficiency of the three-dimensional reconstruction proposed in this paper was approximately three times higher than that of the global match algorithm. Furthermore, the comparison of the manual measurement data of the tree crown width and height with the data from point clouds showed that the accuracy of the point cloud data obtained by this algorithm can meet the needs of large-scale forestry investigations.

Although the algorithm in this paper showed considerable improvement in improving matching efficiency compared with the global match algorithms, there is still room for improvement in fusion feature generation and matching number selection for different sites. Subsequent research will focus on the similarities among the tilted images and the topological relationship between images to further improve the performance of the algorithm, constrain the image to be matched by various means, ensure the accuracy of the generated point cloud and simultaneously improve the algorithm performance.

Acknowledgments

We would like to acknowledge the financial support from National Social Science Foundation as well as assistance from the School of technology of Beijing Forestry University.

This research was funded by National Key Research and Development Plan, grant number 2018YFC0507102-02 and Beijing Natural Science Foundation Project, grant number 6184044 and Huai’an Natural Science Foundation Project, grant number HAB201841 and Jiangsu Province 333 Engineering Research Funding Project , grant number BRA2018350 and Jiangsu Province Professional Leader Training Project, grant number 2019GRGDYX088

REFERENCES

[1] A. Kangas, T. Gobakken, S. Puliti, M. Hauglin, E. Næsset, "Value of airborne laser scanning and digital aerial photogrammetry data in forest decision making", in Silva Fennica, **vol. 52**, pp. 1-19, 2018.

[2] R. Szeliski, "Structure from motion", in Near Eastern Archaeology, **vol. 77**, no. 3, pp. 187-191, 2017.

[3] E.H. Soulaiman, M. Mostafa, E.A. Nabil, S. Abderrahim, S. Khalid, "3D reconstruction system based on incremental structure from motion using a camera with varying parameters", in The Visual Computer, **vol. 34**, no. 10, pp. 1443-1460, 2018.

[4] S. Fuhrmann, F. Langguth, N. Moehrle, M. Waechter, M. Goesele, "MVE-an image-based reconstruction environment", in Comput. Gr., 2015, **vol. 53**, 44-53.

[5] M. Alonzo, H.E. Andersen, D. Morton, B. Cook, "Quantifying boreal forest structure and composition using UAV structure from motion", in Forests. 2018, **vol. 9**, p. 119.

[6] J.P. LI, T. Jiang, D. Xiao, et al., "On diagram based three-dimensional reconstruction of UAV image", in Optics and Precision Engineering, **vol. 24**, no. 6, pp. 1501-1509, 2016.

[7] M.J. Westobya, J, Brasingtonb, N.F. Glassera, M.J. Hambreya, J.M. Reynolds, "Structure from Motion photograametry: A low-cost, effective tool for geoscience applications", in Geomorphology, **vol. 179**, pp. 300-314, 2012.

[8] N. Snavely, Scene Reconstruction and Visualization from Internet Photo Collections. Washington: University of Washington, 2008.

[9] V.H. Hiep, R. Keriven, P. Labatut, J.P. Pons, "Towards high-resolution large-scale multi-view stereo", IEEE, 2009, pp.1430-1437.

[10] Y. Furukawa, J. Ponce, "Accurate, dense, and robust multi-view stereopsis", in IEEE Transactions on Pattern Analysis and Machine Intelligence, **vol. 32**, no. 8, pp. 1362-1376, 2010.

[11] N. Snavely, S.M. Seitz, R. Szeliski, "Skeletal graphs for efficient structure from motion", CVPR, 2008, pp. 1-8.

[12] J. Heinly, J.L. Schonberger, E. Dunn, J.M. Frahm, "Reconstructing the world in six days", IEEE, 2015, pp. 3287-3295.

[13] M. Farenzena, A. Fusiello, R. Gherardi, "Structure-and-motion pipeline on a hierarchical clustertree", IEEE, 2009, pp. 1489-1496.

[14] H.K. AliAkbarpour, G.S. Palaniappan, "Fast structure from motion for sequential and wide area motion imagery", IEEE, 2015, pp. 1086-1093.

[15] K. Sun, W.B. Tao, "A center-driven image set partition algorithm for efficient structure from motion", in Information Sciences, **vol. 479**, pp. 101-115, 2019.

[16] Z.H. Xu, L. Wu, J. Liu, Y.L. Shen, F.S. Li, R. Wang, "Modification of SfM algorithm referring to image topology and its application in 3-dimension reconstruction of disaster area", in Geomatics and Information Science of Wuhan University, **vol. 40**, pp. 599-606, 2015.

[17] Z.X. Song, M.H. Zhang, "3D Reconstruction on Unmanned Aerial Video by Using Patch Clustering Matching Method", in Journal of Northwestern Polytechnical University, **vol. 34**, pp. 731-737, 2016.

[18] H.F. Zhuang, K.Z. Deng, H.D. FAN, "SAR Images Unsupervised Change Detection Based on Com bination of Texture Feature Vector with Maximum Entropy Principle", in Acta Geodaetica et Cartographica Sinica, **vol. 45**, pp. 339-346, 2016.

[19] A. Andoni, P. Indyk, T. Laarhoven, et al., "Practical and optimal LSH for angular distance", in Proceedings of the 28th International Conference on Neural Information Processing Systems. MIT Press, 2015, pp. 1225-1233.

[20] M. Raginsky, S. Lazebnik, “Locality-sensitive binary codes from shift-invariant kernels”, The Neural Information Processing Systems, 2009, pp. 1509-1517.
[21] S. Noah, M. Steven, S. Richard, “Photo Tourism: Exploring image collections in 3D”, in ACM Transactions on Graphics, **vol. 25**, no. 3, pp. 835-846, 2006.
[22] C.C. Wu, “VisualSFM: A Visual Structure from Motion System”, http://ccwu.me/vsfm/, 2011.