

Received January 8, 2020, accepted January 17, 2020, date of publication January 21, 2020, date of current version January 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968335

Spatial Coordinates Correction Based on Multi-Sensor Low-Altitude Remote Sensing Image Registration for Monitoring Forest Dynamics

RUI YU^{1,3,4}, MINGHAO LYU^{3,5}, JIAHUI LU⁶, YANG YANG^{1,3,4}, GUOCHUN SHEN^{3,6}, AND FEI LI²

¹School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China

²Department of Scientific Research, Yunnan Normal University, Kunming 650500, China

³Engineering Research Center of GIS Technology in Western China, Ministry of Education of China, Yunnan Normal University, Kunming 650500, China

⁴Laboratory of Pattern Recognition and Artificial Intelligence, Yunnan Normal University, Kunming 650500, China

⁵School of Forestry, Northeast Forestry University, Harbin 150036, China

⁶Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China

Corresponding authors: Yang Yang (yyang_ynu@163.com), Guochun Shen (gcshen@des.ecnu.edu.cn), and Fei Li (lifei@ynnu.edu.cn)

This work was supported in part by the National Nature Science Foundation of China under Grant 41661080, and in part by the Yunnan Ten-thousand Talents Program.

ABSTRACT Tree species diversity plays a significant role in our ecosystem. In order to monitor forest dynamics, hyperspectral remote sensing equipped on a small unmanned aerial vehicle (UAV) is commonly applied, such as individual tree detection and classification. However, low resolution, positioning errors and the imaging perspective of small UAV affected by wind speed/direction, complex terrain, battery capacity, aircraft posture, flying height and other human factors result in relatively large positional errors (i.e., GPS errors) in such hyperspectral images, and the precise forest dynamics monitoring is limited, especially in spatial analysis. In order to reduce the positional errors of hyperspectral images captured from a small UAV and provide a precise forest dynamics monitoring, we present a novel spatial coordinates correction approach by registering low-altitude UAV visible light and hyperspectral images. The proposed method first employs visible light images and ground control points to stitch a geographic coordinate system as our groundtruth. Hyperspectral images (UHI) are then registered onto the stitched visible light image (UVI) via a novel image registration method. Finally, spatial coordinates of the registered hyperspectral images are updated by using the aforementioned groundtruth. Extensive experiments on image registration and spatial coordinates correction demonstrate the favorable performance of our method. Compared against four state-of-the-art registration methods, our method shows the best registration performance, and the positional errors of hyperspectral images are significantly reduced. Such accuracy is considered very high in this research.

INDEX TERMS Forest dynamics, small unmanned aerial vehicle, multi-sensor image registration, spatial coordinates correction, hyperspectral images.

I. INTRODUCTION

Forest management can effectively change the structure of forest habitats and affect their biodiversity [1]. The forest species richness is one of the important indicators in forest ecosystem services. Meanwhile, forest can strongly influence the urban physical/biological environment by moderating climate, conserving energy, improving urban air quality,

The associate editor coordinating the review of this manuscript and approving it for publication was Stefania Bonafoni¹.

controlling rainfall runoff and flooding, reducing noise levels and sheltering wildlife [2]. Precise evaluation of ecosystem services in forest relies on accurate information of species diversity and spatial distribution of dominant and rare species. However, human activity, soil erosion, pests and natural disaster are causing a drastic decrease in tree species diversity, which requires long-term research and monitoring. Since traditional manual survey based on field inventory work is time-consuming and labor-intensive, there is an urgent need to monitor forest and its dynamics different temporal scales

using a light-weight and low-cost dynamic monitoring technology.

With the technological advances of aerial remote sensing, using aerial images is one of the effective ways for dynamic monitoring of tree species diversity [3]–[7]. Compared with the use of aircrafts or satellites, small UAV remotely information can be cheaper and faster acquired in especially if a higher temporal resolution (even a few centimeters) is required [8], [9]. Therefore, due to the urgent need for very high spatial resolution, it is more common to use a small UAV to carry a complementary metal-oxide semiconductor (CMOS) /hyperspectral camera for monitoring species diversity. Recently, the use of UAV for civilian applications has emerged as an attractive and flexible option for the monitoring of various aspects of forestry and environment [10], [11]. For example, Hassaan *et al.* [12] described an approach based on visible light imagery of UAV to monitor and count trees in urban environment, solving the problem of forest degradation and deforestation. Zhang *et al.* [13] proposed a UAV images registration with large viewpoint, providing an accurate mapping between different viewpoint images for monitoring forest land. Song *et al.* [14], [15] presented a small UAV-based multi-temporal change detection method to monitor cultivated land cover changes in mountainous terrain. Due to hyperspectral remote sensing data obtaining rich spectral information of vegetation, tree species classification is usually based on multispectral/hyperspectral data [16]–[18]. In [19], a method for the precise classification of crops is proposed by using spectral-spatial-location fusion based on conditional random fields (SSLF-CRF) for UAV-borne hyperspectral remote sensing imagery. Khaliq *et al.* [20] provided a detailed analysis and comparison of vineyards MSI, provided by a decametric resolution satellite and low altitude UAV platforms. Rizeei and Pradhan [21] proposed a method of orthorectifying VHR WorldView-3 images by integrating light detection and ranging (LiDAR) data to enhance the urban mapping accuracy.

Compared with satellite and other aerial remote sensing, using small UAV for monitoring tree species diversity has strong flexibility, high efficiency, easy operation, low cost and other advantages. However, issues like the following still exist, hindering the further study: (i) Airborne pushbroom hyperspectral imager has small imaging range and low resolution compared to airborne visible light camera, which inevitably leads to low stitching efficiency, high cost and large geometric error in final result. (ii) Since the imaging perspective is vulnerable to wind speed/direction, complex terrain, battery capacity, aircraft posture, flying height and other human factors, images of the same scene captured by small UAV are always accompanied by multi-viewpoint, which are directly used for stitching to cause stitched image stretching and geometric distortion. (iii) Some small drones can carry direct georeferencing sensors [8], [22]. However, for economy of weight and cost, small UAV hyperspectral imaging system often can not carry position and orientation system with high-precision and high sampling frequency.

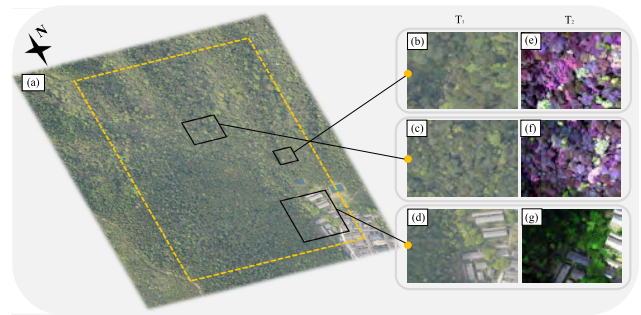


FIGURE 1. The UAV survey in a 20 ha plot in an evergreen broad-leaved forest. (a) a stitched visible light image from data set (I). (b)-(d) different regions from (a). (e)-(g) the hyperspectral images from data set (II), which correspond to (b)-(d), respectively.

(iv) The high-overlap of hyperspectral data shooting easily leads to inefficient work. Especially the hyperspectrum is also susceptible to the weather, the work time is very precious. High-overlap collection is extremely time-consuming and inefficient in mountain forests. Moreover, as the positional error of hyperspectral image is mainly caused by the eccentric angle between the inertial measurement unit (IMU) system of the UAV and the spectral sensor in the UAV pushbroom hyperspectral image [23], [24], the positioning of a single hyperspectral image is inaccurate. The presence of these cumulative errors cause large positional errors, which impedes the elaborate spatial analysis in forest management, such as location-based management.

The above issues have led to the fact that the hyperspectral images captured by small UAV can hardly provide a precise forest dynamics monitoring. Hence, we need a spatial coordinates correction approach by registering low-altitude UAV visible light and hyperspectral images to reduce the positional errors of hyperspectral images. To further facilitate the understanding of the problem of UAV survey, a representative example is provided in Fig 1. In this example, the visible UAV image of the study area in (a) is composed of about 10 visible light images, but it takes about 100 shots with a hyperspectral camera. Meanwhile, the spectral imaging system of small UAV has its own errors and low resolution, which greatly increases the offset of overall position error after plenty of the hyperspectral images are stitched.

A novel approach is proposed to precisely align the low-altitude visible light and hyperspectral images into geographic coordinate system to reduce the positional errors of hyperspectral images. The proposed method first employs visible images and ground control points to stitch a geographic coordinate system as our groundtruth. Then, hyperspectral images are registered onto the stitched visible image via a novel image registration method. Finally, spatial coordinates of the registered hyperspectral images are updated to the proper positions by using the aforementioned groundtruth. This study can greatly improve the efficiency of hyperspectral information acquisition and avoid the traditional high-overlap data acquisition.



FIGURE 2. Location of Chenshan Botanical Garden.

The rest of the paper is organized as follows: Section II introduces the study area and data; Section III develops the main ideas of this work and discuss the computational complexity of the proposed method; Experiments are conducted over remote sensing data sets to comprehensively evaluate the proposed method in Section IV; In Section V, we conclude this work with a discussion.

II. STUDY AREA AND DATA

The study area was conducted in a 1300 m^2 forest plot in Chenshan Botanical Garden (Longitude range: 31°073'N - 31°074'N; Longitude range: 121°185'E - 121°186'E;), located in Songjiang District, Shanghai city, in Eastern China (see Fig 2). The area has a typical subtropical monsoon climate with a mean temperature of 28.1°C and 4.2°C in the warmest and coldest months, respectively. The vegetation is characterized as a subtropical evergreen broadleaf forest.

The image dataset (I) contains a total of 20 visible light images acquired by a small UAV (UVI), the DJI Phantom 4 Pro (DJI, Shenzhen, China, store homepage: [25]) with DJI Zenmuse X5S camera (4/3-inch CMOS image sensor with 20 million pixels, Lens FOV about 84°8.8 mm/ 24 mm), basically maintained the same height (11m-80m). Since the images of the dataset (I) are stitched, image resolutions range from 3000×3000 to 4000×4000 pixels. The image of dataset (I) has been corrected and stitched by the photogrammetry software PHOTOMOD so that it has high-precision spatial geographic coordinates. The stitched images are conducted by the bundle adjustment. Thus, the geographic coordinate system of stitched visible light image can be regarded as our groundtruth.

The image dataset (II) contains a total of 400 hyperspectral images acquired by a small UAV (UHI), the DJI M600 pro (DJI, Shenzhen, China, store homepage: [26]) with Gaia sky-mini pushbroom hyperspectral imaging system (charge-coupled device (CCD) Sony ICX674 sensor with a spectrum resolution of approximately 3±0.5 nm, Dualix instruments Ltd., Sichuan, China, store homepage: [27]), basically maintained the same height (20m-80m),

TABLE 1. Technical details of hyperspectral flight mission.

Item	Detail
Platform	DJI M600 pro
Sensor	Gaia Sky-mini2
Number of Channels	176
Spectral Band Range(nm)	392.7-1009.4
Relative Flying Height(m)	61
Ground Sampling Distance(cm)	3
Flight Speed(m/s)	5
Forward Overlap(%)	75
Side Overlap(%)	70

and image resolutions range ranges from 1000×1200 to 1920×1400 pixels. Each of the hyperspectral images have only one geographic coordinate.

In this paper, we adopt Trimble R4-3 to collect ground control points used in the bundle adjustment. A single image contains inaccurate exterior orientation. The calibration parameters are obtained by the camera manufacturer's calibration, and the calibration process is completed by software. Table 1 gives the technical details of hyperspectral flight mission.

III. METHODOLOGY

The main process of the proposed method includes three major steps: 1) geographic coordinate system determination, 2) image registration, and 3) spatial coordinates correction as shown in Fig 3. In this section, we first introduce the proposed method followed by analyzing the computational complexity and discussing the implementation details.

A. GEOGRAPHIC COORDINATE SYSTEM DETERMINATION

We utilize high resolution UVI and ground control points to stitch a geographic coordinate system as our groundtruth. All stitched results are corrected and stitched by the photogrammetry software PHOTOMOD. Owing to the different imaging scale/height between the CMOS sensor and the hyperspectral sensor, the different region of the UHI needs to be determined on the UVI by the geographic coordinates before registration. Given that the geographic coordinates of the UVI and the UHI, the goal is to generated a suitable region that is mapped from the UHI to UVI.

We select the appropriate regions based on the hyperspectral image in the stitched visible light images (UVI). For each UVI region (UVR) I_v^r , we adopt the parameterizations of the 4 variables following

$$\begin{aligned}
 c_{la} &= \frac{c}{v_{la}^{max} - v_{la}^{min}}(v_{la}^{max} - x_{la}), \\
 c_{lo} &= \frac{r}{v_{lo}^{max} - v_{lo}^{min}}(x_{lo} - v_{lo}^{min}), \\
 p_h &= \frac{h_v}{h_h} \times c_h, \\
 p_w &= \frac{h_v}{h_h} \times r_h,
 \end{aligned} \tag{1}$$

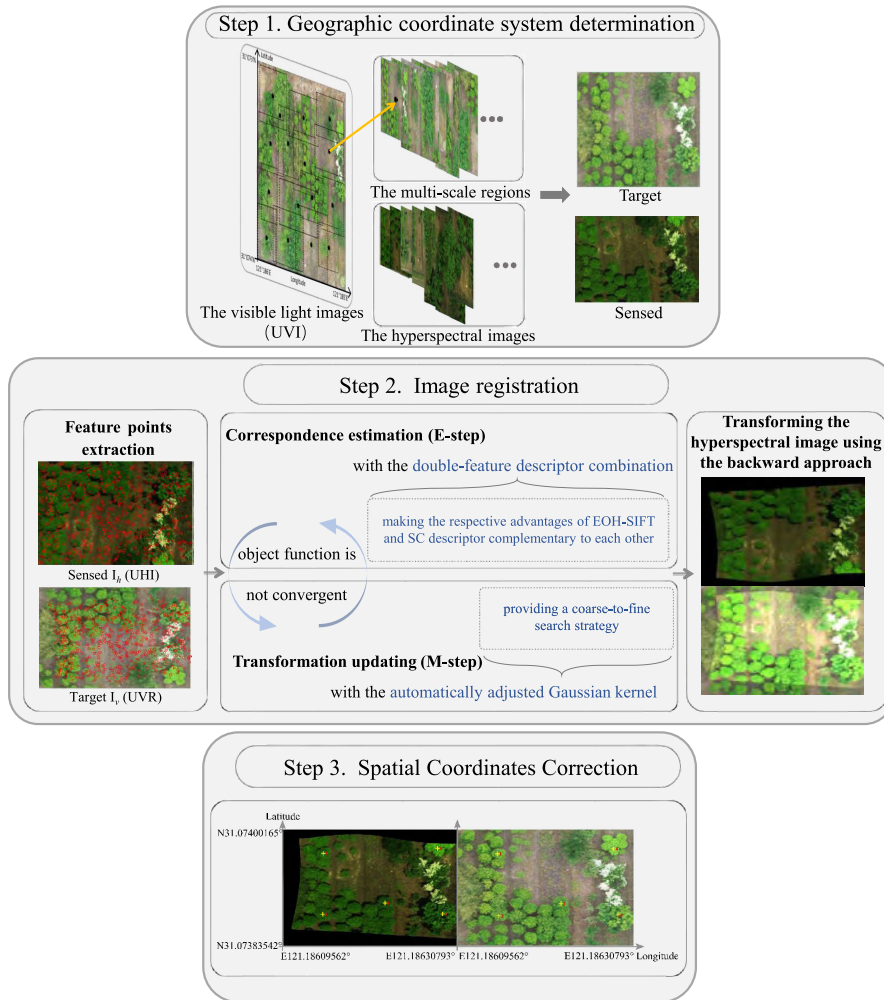


FIGURE 3. The main process of our method. In the step 1, black dotted bordered rectangles denote the selected regions of interest from the visible light image, and the blackspot denote the center of an image, corresponding to a hyperspectral image in the second column. In the step 2, red circles denote the extracted feature points.

where c_{la} , c_{lo} , p_h , and p_w denote the regions' central coordinates and its height and width, and the $c \times r$ visible stitched image mainly has two geographic coordinates (v_{la}^{max} , v_{lo}^{max}) and (v_{la}^{min} , v_{lo}^{min}), the UHI' GPS coordinate is (x_{la}, x_{lo}) . h_v and h_h represent the shooting height of the UVI and the UHI, respectively, and c_h and r_h are the width and height of the UHI.

B. IMAGE REGISTRATION

After we obtain a geographic coordinate system as our groundtruth, the UHI and UHR are placed in image registration. The goal of image registration is to transform a hyperspectral image I_h (UHI) so that it is aligned to a target image I_v' (UVR). Thus, more accurate geographic coordinates are aligned on the UHI to locate the tree species.

1) MAIN PROCESS

The main process of the proposed feature-based image registration method normally has a three-steps process as follows:

- Feature points extraction. In the feature-based methods, features extraction plays an essential role in point set registration. The feature points extracted from multi-sensor small UAV images with severe appearance changes often suffer from the problems of insufficient feature points and high outlier ratio. Therefore, the Edge Oriented Histogram descriptor based on a (scale-invariant feature transform) SIFT-like scale space (EOH-SIFT) [28] (see Section III-B2) is employed for the feature point sets extraction. A sensed feature point set $H = \{h_i\}_{i=1}^N$ and a target feature point set $V = \{v_j\}_{j=1}^N$ are extracted from I_h and I_v' , respectively.
- Feature point sets registration. In this work, we employ the expectation maximization (EM) [29] algorithm to finish the image registration process, which alternates between two steps:
 - (a) Expectation step (E-step): guessing the values of parameters ("old" parameter values estimated in previous iteration) used to compute posteriori

probability distributions of mixture components based Bayes' rule;

- (b) Maximization step (M-step): founding the “new” parameter values via minimizing the expectation of the complete negative log-likelihood function.

These two steps of EM algorithm correspond to correspondence estimation and transformation updating as follows:

Correspondence Estimation: In this work, the registration of H and V is considered as a Gaussian mixture model (GMM) probability density estimation problem. The GMM is constructed by N Gaussian components, and let h_i be the centroid of the i^{th} component, v_j the j^{th} data. Thus, the probability density function of the GMM is obtained as $p(h_i) = \frac{\kappa}{N} + (1 - \kappa) \sum_{j=1}^N \frac{r_{ij}}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2} \|h_i - v_i\|^2)$ where the priors r_{ij} are non-negative quantities for which $\sum_{i=1}^N r_{ij} = 1$, which weighting for each Gaussian component. $\frac{1}{N}$ is an additional uniform distribution with $0 \leq \kappa \leq 1$, for the outliers and noise.

In order to obtain a more reliable correspondence estimation via feature complementarity, a new cost matrix \mathcal{D}^c (see Section III-B2) is used to determine the weighting matrix $R_{N \times N} = \{r_{ij}\}_{i=1, j=1}^{N, N}$ of the GMM.

The weighting matrix R based on \mathcal{D}^c is solved as a linear assignment problem: $\mathcal{D}^c(R) = \sum_{i=1}^N \sum_{j=1}^N r_{ij} \mathcal{D}_{ji}^c$, by the Jonker-Volgenant algorithm [30]. The permutation matrix R is constructed by $r_{ij} \in \{0, 1\}$, and R still satisfies $\sum_{i=1}^N r_{ij} = 1$.

Based on the Bayes' rule, the posteriori probability matrix P (E-step) of GMM is computed as follow

$$p_{ij} = \frac{r_{ij} \exp(-\frac{\|h_i - v_i\|^2}{2\sigma^2})}{\frac{\kappa}{1-\kappa} \frac{1}{N} + \sum_{k=1}^N r_{ki} \exp(-\frac{\|h_i - v_i\|^2}{2\sigma^2})}, \quad (2)$$

and $P_{N \times N}$ is considered as a probability correspondence matrix. Then we can obtain the putative target point set \tilde{V}_p for updating H by $\tilde{V}_p = PV$.

Transformation updating: The non-rigid feature point set transformation function is formed as follow

$$\mathcal{T}(H, \Psi) = H + \Gamma\Psi, \quad (3)$$

where Γ is an automatically adjusted Gaussian kernel (see Section III-B3), and the updating is to estimate a suitable coefficient matrix Ψ for the transformation \mathcal{T} . Based on the probability density function of the GMM, a reliable displacement direction will produce a larger expectation of probabilities. Hence, the solution of the transformation updating is detected by maximizing a likelihood function that is formed as $\prod_{j=1}^N p(r_j)$, or equivalent to minimizing the negative log-likelihood function, which is formed as

$$\mathcal{Q}(\kappa, \Psi, \sigma^2) = - \sum_{j=1}^N \log[(1 - \kappa) \sum_{i=1}^N r_{ij} \frac{\exp(-\frac{\|h_i - v_i\|^2}{2\sigma^2})}{2\pi\sigma^2} + \frac{\kappa}{N}] + \mathcal{L}, \quad (4)$$

where $\mathcal{L} = \frac{\alpha}{2} \text{tr}(\Psi^T \Gamma \Psi)$ represents a regularization term based on Motion Coherence Theory (MCT) [31], and $\text{tr}(\cdot)$ denotes the trace operate.

Then the optimal parameter Ψ is estimated by

$$\Psi = \arg \min_{\Psi} \mathcal{Q}(\kappa, \Psi, \sigma^2). \quad (5)$$

The solution of the parameter estimation is given in Appendix A. The locations of \hat{H} are then updated by Equation 3.

- Image transformation.

The thin plate spline is usually used in computer vision and image analysis [32], [42], [44], [45], particularly in non-rigid transformation model. In order to solve image geometric distortion and non-rigid distortions, we have utilized TPS to simulate a wider transformation. We apply the initial H and its final location \hat{H} as a corresponding set \mathfrak{R} to build the thin plate spline (TPS) based image transformation using the backward approach. The image I_h is transformed by:

$$Y^{TPS} = \begin{pmatrix} \mathcal{G} & \Phi \\ \Phi^T & \mathbf{O} \end{pmatrix}^{-1} (H \ \mathbf{0} \ \mathbf{0} \ \mathbf{0})^T, \quad (6)$$

where Y^{TPS} is of size $(N + 3) \times 3$, \mathbf{O} is a 3×3 matrix of zeros, Φ is the $N \times 3$ matrix with the i^{th} denoting $(1, t_i)$, and the $N \times N$ TPS kernel $\mathcal{G}_{ij} = \|t_i - t_j\|^2 \log \|t_i - t_j\|$, where t_i and t_j indicate the coordinates of \hat{H} . Then a regular grid $\Theta_{Z \times 2}^t = \{\theta_z^t\}_{z=1}^Z$ is obtained by a pixel-by-pixel indexing process on the sensed image I_v , where $Z = x(I_h) \times y(I_h)$ with $x(\cdot) \times y(\cdot)$ denotes the size of the image. Let grid Θ^t be the feature points from the reference image, the transformed grid is obtained by first computing $\hat{\Theta}_{Z \times 3}^t = (\overline{\mathcal{G}} \ \overline{\Phi}) Y^{TPS}$, then restoring the dimension of the grid to 2 by $\hat{\Theta}^t \leftarrow (\hat{\Theta}_{(\cdot, 1)}^t \ \hat{\Theta}_{(\cdot, 2)}^t)$, where the $Z \times N$ kernel $\overline{\mathcal{G}}_{ij} = \|\theta_i^t - t_j\|^2 \log \|\theta_i^t - t_j\|$, $\overline{\Phi}$ is the $Z \times 3$ matrix with the z^{th} row denotes $(1, \theta_z^t)$ and $\hat{\Theta}_{(\cdot, i)}^t$ denotes the i^{th} column of $\hat{\Theta}^t$. Let Θ be the grid obtained on I_h , we have

$$\hat{\Theta} = \hat{\Theta}^t \cap \Theta. \quad (7)$$

Finally, the transformed image I^t is obtained by getting intensities from the hyperspectral image I_h based on $\hat{\Theta}$, and setting the rest of pixels to black. Note that the bicubic interpolation is used to improve the smoothness of I^t , to be more precisely, the intensities of each pixel in I^t is determined by summing the weighted neighboring pixel intensities within a 4×4 window.

2) DOUBLE-FEATURE DESCRIPTOR COMBINATION

The double-feature descriptor combination makes the respective advantages of EOH-SIFT and shape context descriptor complementary to each other.

The EOH-SIFT algorithm [28] is a feature point descriptor for matching feature sets on visible spectrum images of the same scene. Firstly, feature sets are detected through

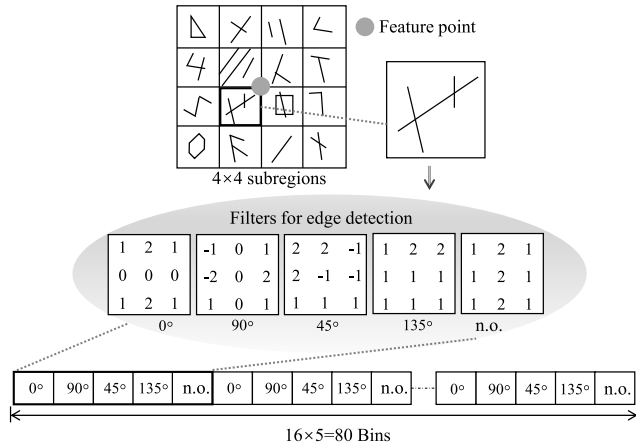


FIGURE 4. The illustration of the EOH-SIFT descriptor [28]. Firstly, the $N \times N$ pixels region centered on a given feature point is obtained. Then, the region is divided into $4 \times 4 = 16$ subregions. Finally, each of these subregions is represented by a contour histogram calculated after the edge histogram descriptor (EHD) [33] of the MPEG-7 standard [34]. The histogram represents the spatial distribution of four directional edges (0, 45, 90, 135 degrees) and one non-directional edge (n.o.).

a SIFT-like based scale space representation, and then an Edge Oriented Histogram (EOH) descriptor is used to represent these feature sets, obtaining the 80-dimensional vector as the EOH-SIFT descriptor, as shown in Fig 4. These EOHs incorporate spatial information from the contours near each feature point and describe the shape and contour of the image, maintaining the invariance of the scale space. Finally, feature sets from multispectral images are matched by finding nearest couples using the information from the descriptor.

Compared with SIFT, EOH-SIFT descriptor improves its robustness via (i) discarding feature point which their subregions without information (i.e., only a few contours are contained in subregions), and (ii) using the scale restriction. The scale restriction process uses the scale difference (SD) of the given pair of feature sets H and V used as follow:

$$SD(H, V) = \varrho_h - \varrho_v, (a < SD < b) \quad (8)$$

where ϱ is the scale of that pyramidal representation that the feature point appears. The sough value of a and b are obtained by (i) computing a histogram of SDs of all matches, and (ii) extracting the peak \overline{SD} in the SDs histogram. They are defined as $a = \overline{SD} - 0.9, b = \overline{SD} + 0.9$.

The shape context descriptor [35] is a classic algorithm, which lets feature points lie within a polar coordinate system with $B_{\mathcal{R}}$ bins in the radial direction and $B_{\mathcal{T}}$ bins in the tangential direction, by centering the coordinate at s_i and r_j . The count of feature points in all bins is denoted by B . The shape context chi-square matrix is defined as follow:

$$SC(h_i, v_j) = \frac{1}{2} \sum_{b=1}^B \frac{[x_i^h(b) - x_j^v(b)]^2}{x_i^h(b) + x_j^v(b)}, \quad (9)$$

where $p_i^h(b)$ and $p_j^v(b)$ denote the number of points in the b^{th} bin of the coordinate centered at h_i and v_j , respectively.

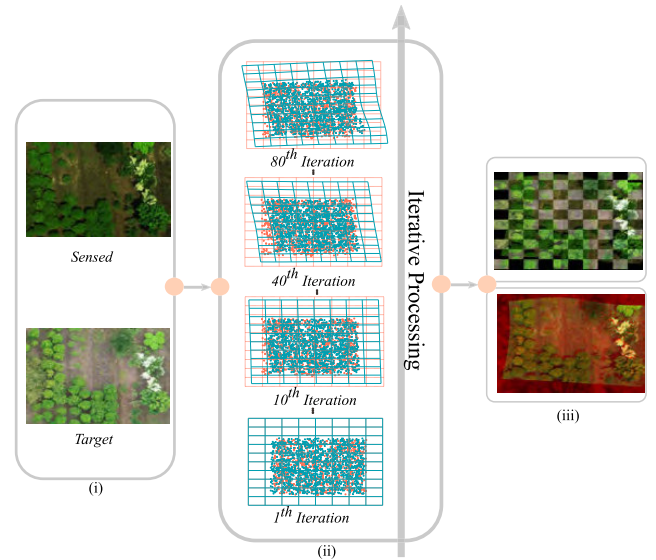


FIGURE 5. The illustration of the proposed automatically adjusted Gaussian kernel processing. (i) A pair of multi-sensor UAV images, where the upper and lower are the sensed and target images. (ii) The block from bottom to top successively shows the iterative process of feature point sets from rigid to non-rigid transformation. The orange and blue dots (crosses) denote inliers (outliers) in two feature point sets, respectively. The orange grids (orange dotted lines) and the blue grids (blue solid lines) denote the original image field and the warped image field, respectively. (iii) For visual comparison, the image registration result (i.e., the transformed image) is shown with its target image by a 10×10 montages in the first row, and the second row shows the overlapping area.

Thus, we combine the aforementioned two features to compute computing a double-feature descriptor combination cost matrix through an element-wise Hadamard product (expressed by \odot)

$$D^c = SD(H, V) \odot SC, \quad (10)$$

In summary, the used double-feature descriptor combines the spectrum information and geometrical structure to improve the description of the feature point set and provides an accurate guiding for feature sets registration.

3) AUTOMATICALLY ADJUSTED GAUSSIAN KERNEL

The automatically adjusted Gaussian kernel is proposed to gradually change the transformation function updating from rigid to non-rigid via controlling the displacement distances of the feature point sets. The idea of this kernel is to play a coarse-to-fine search strategy.

The non-rigid transformation function is reproduced in the a specific functional space \mathcal{H} , namely a vector-valued Reproducing Kernel Hilbert Space (RKHS) [36], which applies the form as $\mathcal{T}(H, \Psi) = H + Dis(H)$ with a displacement function Dis . Myronenko and Song [37] have used the Gaussian radius basis function (GRBF) to define the RKHS by a reproducing Gaussian kernel $\Gamma_{N \times N}$ with a coefficient constant $W_{N \times D}^c$. The displacement function is formed as $Dis(H) = \sum_{i=1}^N \Gamma(h(\cdot, \cdot), h_i) w_{(i, \cdot)}^c$, where the Gaussian kernel is defined by $\Gamma(h_j, h_i) = \exp(-\frac{\|h_i - h_j\|^2}{2\gamma})$, where the constant γ controls

Algorithm 1 The proposed method

```

input : The UVI  $I_v$  and the UHI  $I_h$ 
1 Determination of regions of interest:
2   Compute the target image region  $I'_v$  using
   Equation (1);
3 end
4 Image registration:
5   Initialise parameter  $w, \alpha, B_{\mathcal{R}}, B_{\mathcal{T}}, \gamma, \eta, \mathcal{G}, \kappa, \sigma^2,$ 
    $\Psi$  and  $iter_{max}$ ;
6   Extract feature set  $H$  and  $V$  from  $I_h$  and  $I'_v$  using
   EOH-SIFT, respectively;
7   Construct the Gaussian kernel  $\Gamma$ ;
8   do
9     E-Step:
10    Compute  $P$  using Equation (2);
11    Compute corresponding target point set
     $\tilde{V}_p$  by  $PV$ ;
12   end
13   M-Step:
14   Update  $\Psi$  using Equation (17);
15   Compute  $\mathcal{T}(H, \Psi)$  using Equation (3)
    for updating  $H$ ;
16   Update  $\kappa$  and  $\sigma^2$ , respectively;
17   end
18   while Equation 16 is not convergent;
19   Annealing  $\gamma \leftarrow t \times \gamma$ ;
20   Compute the transformed image  $I^t$  using
   Equation (7).
21 end
22 Spatial coordinates correction:
23   Compute the corrected spatial coordinates of the
   UHI using Equation (12);
24 end
output :  $\mathcal{T}, I^t, (v'_{la}, v'_{lo})$ 
    
```

the strength of displacement. The weight term W^c holds the displacement direction of V .

Based on the GRBF based non-rigid transformation function, the automatically adjusted Gaussian kernel is designed by adjusting the γ via an annealing scheme. The temperature of the annealing is formed as follow:

$$t = \eta \cdot \exp(\sigma^2), \tag{11}$$

where the constant η denotes the final temperature, σ^2 is the decreasing notice radius used in the correspondence estimation step. Thus, the displacement γ is renewed by $\gamma \leftarrow t \times \gamma$. Fig 5 illustrates the proposed automatically adjusted Gaussian kernel processing with the recovered warping grids.

Thereinto, η has a greater value, the changes of temperature is more drastic and the maximum of temperature is higher. In other words, the strength of displacement is stronger and its duration is longer in the earlier stage when η is greater.

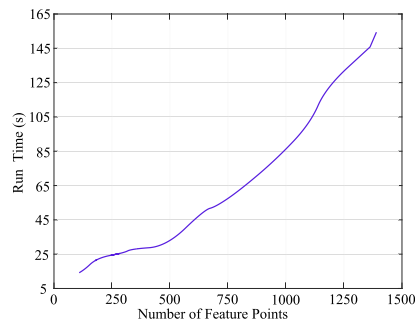


FIGURE 6. The run time of the proposed method with different number of feature points.

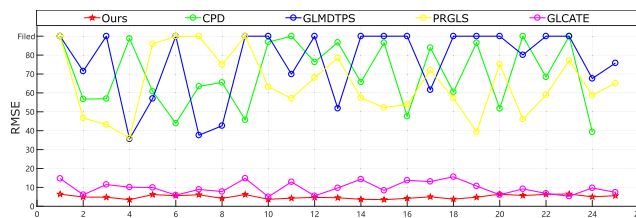


FIGURE 7. RMSEs of each method for the image registration.

TABLE 2. Quantitative comparison on image registration measured using the mean RMSE, MAE and MAD are carried out. Bold fonts indicate the best results. All units are in pixel.

Method	RMSE	MAE	MAD
OURS	5.0235	6.0028	1.0728
CPD	82.0995	92.1365	40.9872
GLMDTPS	84.8720	95.7640	44.3518
PRGLS	66.6927	66.1968	27.8758
GLCATE	9.7241	78.4462	21.0610

C. SPATIAL COORDINATES UPDATE

After image transformation, the spatial coordinates of the UVR can be accurately superimposed on the UHI. Then, for each pixel of the UHI, we can obtain its geographic coordinates (v'_{la}, v'_{lo})

$$\begin{aligned}
 v'_{la} &= v_{la}^{max} - c_{la} \frac{v_{la}^{max} - v_{la}^{min}}{c}, \\
 v'_{lo} &= r_{lo} \frac{v_{lo}^{max} - v_{lo}^{min}}{r} + v_{lo}^{min}
 \end{aligned} \tag{12}$$

where c_{la}, c_{lo} denote the position of pixel in the UHI, and the $c \times r$ visible stitched image mainly has two geographic coordinates $(v_{la}^{max}, v_{lo}^{max})$ and $(v_{la}^{min}, v_{lo}^{min})$.

D. IMPLEMENTATION DETAILS

1) PARAMETER SETTINGS

For evaluating the proposed method, eight groups of parameters are used in our method.

- w is the window size for every feature point in the EOH-SIFT descriptor, set to 86;

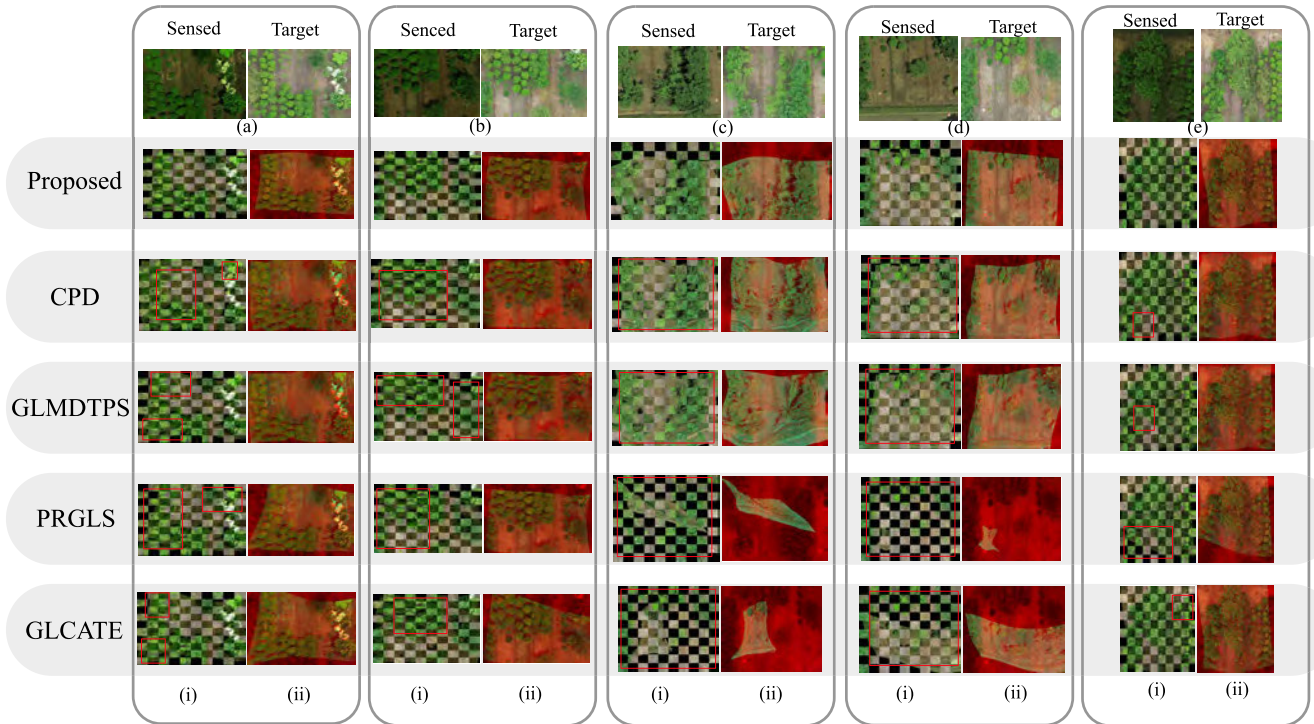


FIGURE 8. Registration examples of the aforementioned methods. Four representative registration examples (a)-(e) and comparisons of our method against CPD [37], GLMDTPS [38], PRGLS [39] and GLCATE [40], on datasets (I) and (II). The first column shows the sensed images and the target images. (i) gives the 10×10 checkerboard for each example, where the registration errors are highlighted using the red rectangle. (ii) gives the overlapping area.

- $B_{\mathcal{R}}$ and $B_{\mathcal{T}}$, the numbers of bins for the shape context descriptor, are set to $B_{\mathcal{R}} = 12$, $B_{\mathcal{T}} = 5$;
- κ is outlier weighting parameter, set to 0.5. It is updated by $\kappa = 1 - N_p/N$.
- σ^2 is the covariance of the double-feature descriptor combination. Initialising and updating the variance σ^2 by

$$\sigma^2 \leftarrow \frac{1}{2N_p} |tr(\tilde{V}^T d(P1)\tilde{V}) - 2tr(\mathcal{T}(\tilde{H}, \Psi)^T P\tilde{H}) + tr(\mathcal{T}(\tilde{H}, \Psi)^T d(P_T 1)\mathcal{T}(\tilde{H}, \Psi))|. \quad (13)$$

- Ψ is the parameter of point set transformation. We initialize the coefficient constant Ψ as a matrix with all zeros;
- α is the weighting parameter of regularization, which is set to 8;
- γ and η are the strengthen constants for the automatically adjusted Gaussian kernel, set as $\gamma=5$ and $\eta=0.8$.
- $iter_{max}$, the max number of iteration, is set to 80.

The pseudo-code of the proposed method is summarized in the Algorithm 1.

2) COMPUTATIONAL COMPLEXITY

The derivative Equation (17) is of $\mathcal{O}(N^3)$ time complexity owing to the existence of the dynamic Gaussian kernel Γ size of $N \times N$. Overall, the time complexity of our method is $\mathcal{O}(N^3)$. For storing kernel Γ , the space complexity of our method is $\mathcal{O}(N^2)$. The run time of the proposed method with different number of feature points is exemplified in Fig 6.

IV. EXPERIMENTS AND RESULTS

A. EXPERIMENTS DESIGN

We design two types of experiments: (i) Quantitative comparison and qualitative demonstration on image registration are carried out on all methods using the root mean square error (RMSE), maximum error (MAE) and median absolute deviation (MAD). CPD (coherent point drift) [37], GLMDTPS (global and local mixture distance with thin plate spline transformation) [38], PRGLS (preserving global and local structures) [39] and GLCATE (global-local correspondence and transformation estimation) [40], four state-of-the-methods, are compared with the proposed method in the following experiments on the dataset (I) and (II). (ii) Since the species coordinate correction is performed by image registration, the quantitative comparison of individual tree species location adopts the actual location. All experiments are implemented in Matlab2017a on a laptop with a 2.6-GHz Intel Core CPU and 8-GB RAM.

B. IMAGE REGISTRATION EVALUATION CRITERIA

The RMSE, MAE and MAD are usually used to measure the image registration error [41]–[46], which are defined as follows:

$$RMSE = \sqrt{\frac{1}{N^L} \sum_{n=1}^{N^L} d(a_n^L, b_n^L)^2},$$

$$MAE = \sqrt{\frac{1}{N^L} \sum_{n=1}^{N^L} (|a_{(n,1)}^L - b_{(n,1)}^L| + |a_{(n,2)}^L - b_{(n,2)}^L|)^2},$$

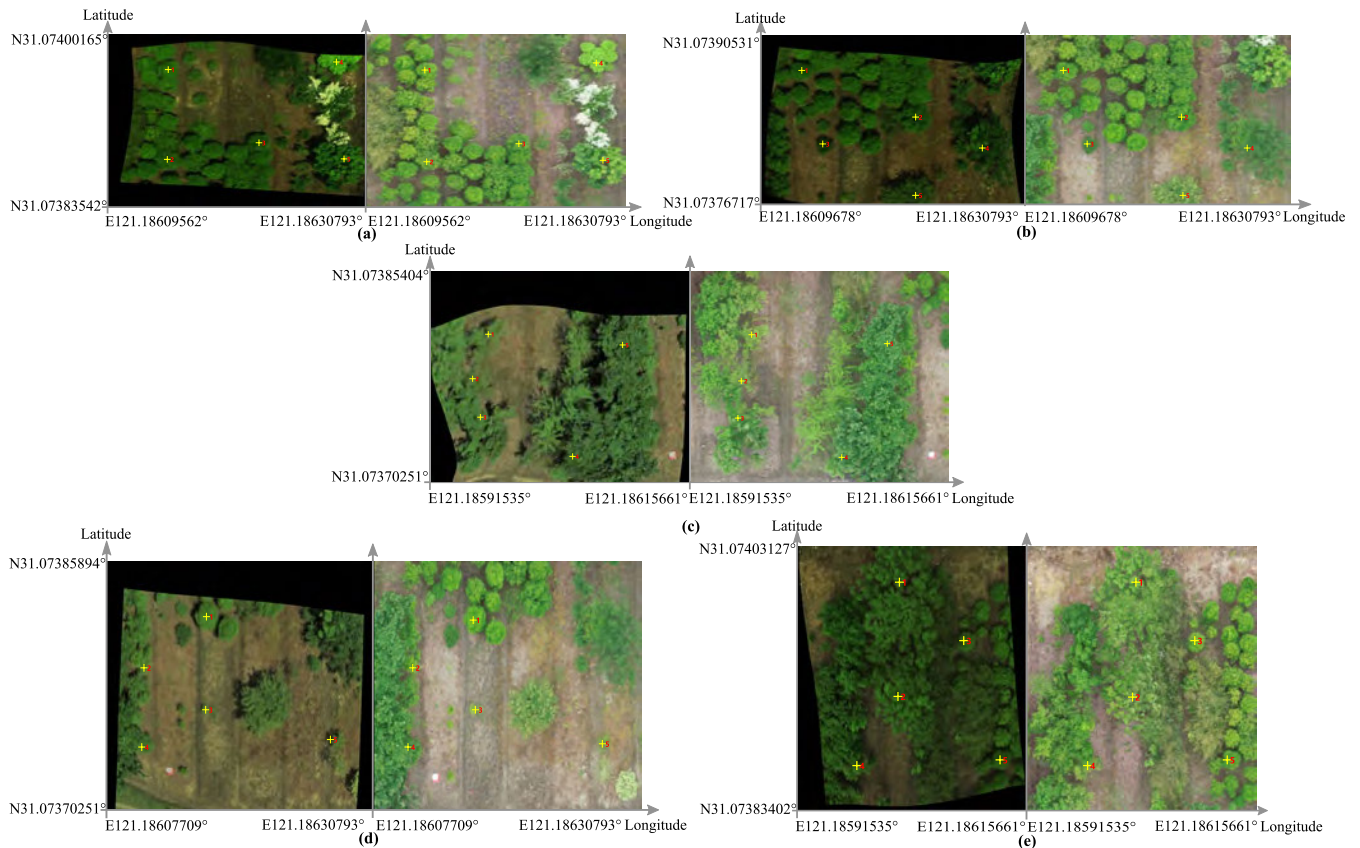


FIGURE 9. Five typical examples of the results for the coordinate optimization. For (a) to (e), the left column gives the transformed image with the manually marked landmarks, and the right column shows the target image with the manually marked landmarks, where the landmarks of the two image correspond to each other. Each of landmarks has the corrected coordinate in the transformed image. The corresponding tree species for each landmark are as follows: *Osmanthus fragrans*: (a)-1~(a)-3, (b)-1, (b)-2, (b)-5, (d)-1~(d)-4, (e)-1, (e)-2 and (e)-5. *Diospyros kaki*: (a)-4. *Magnolia*: (a)-5, (c)-3, (e)-4. *Viburnum propinquum*: (b)-3. *Persimmon*: (b)-4. *Cerasus campanulata*: (c)-1, (c)-2. *Gleditsia vestita*: (c)-4, (c)-5. *Callicarpa candicans*: (d)-5. *Choerospondias axillaris*: (e)-1, (e)-2.

TABLE 3. The results of comparison between the single and the actual location.

	UHI	Actual location (Groundtruth)	Errors
(a)	31.07387910°N, 121.18619574°E	31.07390949°N, 121.18621526°E	3.86m
(b)	31.07384813°N, 121.18619574°E	31.07381191°N, 121.18621283°E	4.34m
(c)	31.07372976°N, 121.18603765°E	31.07375461°N, 121.18605103°E	3.04m
(d)	31.07372868°N, 121.18619722°E	31.07374611°N, 121.18619261°E	1.99m
(e)	31.07395449°N, 121.18603987°E	31.07394396°N, 121.18603259°E	1.36m
		Average errors	2.92m

TABLE 4. The results of comparison between the corrected location and the actual location.

	Actual location	Corrected location	Errors	Correction ratio
(a)	31.07390949°N, 121.18621526°E	31.07391935°N, 121.18622107°E	1.22m	68.19%
(b)	31.07381191°N, 121.18621283°E	31.07382464°N, 121.18620988°E	1.44m	66.71%
(c)	31.07375461°N, 121.18605103°E	31.07376586°N, 121.18604968°E	1.25m	58.74%
(d)	31.07374611°N, 121.18619261°E	31.07375312°N, 121.18619984°E	1.04m	47.28%
(e)	31.07394396°N, 121.18603259°E	31.07395201°N, 121.18603301°E	0.89m	34.64%
		Average errors	1.17m	55.11%

$$MAD = \mathcal{M}(|d(a_n^L, b_n^L) - \mathcal{M}(d(a_n^L, b_n^L))|), \quad (14)$$

where a_n^L and b_n^L are the n^{th} pair of corresponding landmarks and (n, x) denotes the coordinate at x^{th} dimension, N^L is

the total number of selected landmarks, and $\mathcal{M}(\cdot)$ returns the median of a vector, and the operator $d(\cdot, \cdot)$ denotes the distance. At least 20 pairs of corresponding points for each image pair (I_v and I_h') are manually marked as landmarks, and

TABLE 5. The results of comparison between the location in UHI and the actual location of individual tree. DE means the deviation errors.

	Location in UHI-tree	Actual location-tree	DE
(a)-1	31.07394798°N,121.18613560°E	31.07396605°N,121.18614503°E	4.03± 2.83m
(a)-2	31.07387050°N,121.18614092°E	31.07388016°N,121.18614657°E	
(a)-3	31.07388591°N,121.18621827°E	31.07389682°N,121.18627787°E	
(a)-4	31.07396014°N,121.18629326°E	31.07396997°N,121.18629326°E	
(a)-5	31.07387444°N,121.18629336°E	31.07388180°N,121.18630060°E	
(b)-1	31.07386121°N,121.18612755°E	31.07387396°N,121.18612920°E	3.48± 2.36m
(b)-2	31.07382725°N,121.18621217°E	31.07383575°N,121.18622609°E	
(b)-3	31.07380326°N,121.18614750°E	31.07381322°N,121.18614966°E	
(b)-4	31.07380850°N,121.18628051°E	31.07383509°N,121.18622764°E	
(b)-5	31.07376645°N,121.18623862°E	31.07377370°N,121.18622802°E	
(c)-1	31.07379915°N,121.18597455°E	31.07381387°N,121.18597209°E	1.70± 0.68m
(c)-2	31.07377248°N,121.18596147°E	31.07378382°N,121.18596321°E	
(c)-3	31.07377018°N,121.18596147°E	31.07378382°N,121.18596013°E	
(c)-4	31.07371050°N,121.18606338°E	31.07372765°N,121.18605431°E	
(c)-5	31.07379106°N,121.18609075°E	31.07380995°N,121.18609446°E	
(d)-1	31.07381255°N,121.18616771°E	31.07382170°N,121.18616819°E	3.70± 2.68m
(d)-2	31.07377572°N,121.18611470°E	31.07379395°N,121.18611569°E	
(d)-3	31.07371080°N,121.18617368°E	31.07376815°N,121.18617205°E	
(d)-4	31.07372954°N,121.18612062°E	31.07374529°N,121.18610951°E	
(d)-5	31.07374008°N,121.18628284°E	31.07374921°N,121.18628438°E	
(e)-1	31.07399115°N,121.18601083°E	31.07401405°N,121.18603193°E	2.49± 0.75m
(e)-2	31.07390413°N,121.18600999°E	31.07392261°N,121.18602922°E	
(e)-3	31.07394340°N,121.18607829°E	31.07396768°N,121.18609214°E	
(e)-4	31.07384784°N,121.18597820°E	31.07386808°N,121.18598251°E	
(e)-5	31.07385775°N,121.18612276°E	31.07387298°N,121.18612688°E	
	Average errors		2.35m

all the landmarks are selected and well distributed at easily identified places.

C. RESULTS OF IMAGE REGISTRATION

For each method, all extracted features points were adopted for image registration. The RMSEs of each method for 25 image pairs are shown in Fig. 7.

The RMSE, MAE and MAD for each method on 25 image pairs are shown in Table 2. Five typical registration examples are shown in Fig 8.

The failed registrations are determined by two ways: (i) the transformed image is so distorted that it can not be recognized artificially, and (ii) the RMSE is larger than 160 pixels. Our method provides the best performances in all image registration experiments, especially when these images with different imaging sensors have large intensity variations. In the case of CPD, this problem is alleviated by rejecting outliers using an uniform distribution. In addition, GLMDTPS performs poorly as it forces one-to-one correspondence, which is vulnerable to the presence of outliers. PRGLS suffers from dubious correspondences generated by similar geometrical neighborhood structures. GLCATE performs relative well

compared with the other three methods, yet it is insensitive to the multi-temporal and changes image pairs.

The major reasons for outperforming the other four methods are that: (i) the proposed method more accurately identified both feature points by using the double-feature descriptor combination; (ii) the identified feature points are used to shape the warping grids approximately within both overlap and non-overlap areas by the coarse to fine transformation.

D. RESULTS OF THE SPATIAL COORDINATES CORRECTION

1) THE SPATIAL COORDINATES CORRECTION OF A SINGLE UHI

We calculate distances between points described in terms of latitude and longitude. When latitude is the same, every 1.0×10^{-5} degree of longitude produces a distance difference of 1 meter. When longitude is the same, every 1.0×10^{-5} degree of latitude produces a distance difference of 1.1 meters. Table 3 shows the error between the geographic coordinates of single and the geographic coordinates of actual location. In this case, large distortion occurs when a large number of hyperspectral images are stitched, and the overall error is larger than the average error.

TABLE 6. The results of comparison between the corrected location and the actual location of individual tree. DE means the deviation errors. CR means the correction ratio.

	Corrected location-tree	Actual location-tree	DE	CR
(a)-1	31.07396324°N,121.18613989°E	31.07396605°N,121.18614503°E		73.78%
(a)-2	31.07388360°N,121.18613947°E	31.07388016°N,121.18614657°E	0.65± 0.13m	35.01%
(a)-3	31.07389682°N,121.18621729°E	31.07389682°N,121.18622494°E		89.64%
(a)-4	31.07397042°N,121.18628780°E	31.07396997°N,121.18629326°E		71.56%
(a)-5	31.07388269°N,121.18629241°E	31.07388180°N,121.18630060°E		27.11%
(b)-1	31.07387696°N,121.18612253°E	31.07387396°N,121.18612920°E		
(b)-2	31.07383264°N,121.18622140°E	31.07383575°N,121.18622609°E	0.52± 0.25m	65.01%
(b)-3	31.07380899°N,121.18614323°E	31.07381322°N,121.18614966°E		31.34%
(b)-4	31.07383986°N,121.18623173°E	31.07383509°N,121.18622764°E		88.74%
(b)-5	31.07377398°N,121.18622501°E	31.07377370°N,121.18622802°E		79.15%
(c)-1	31.07380904°N,121.18596770°E	31.07381387°N,121.18597209°E		
(c)-2	31.07377669°N,121.18595958°E	31.07378382°N,121.18596321°E	0.65± 0.2m	32.51%
(c)-3	31.07377916°N,121.18595895°E	31.07378382°N,121.18596013°E		65.77%
(c)-4	31.07372359°N,121.18605376°E	31.07372765°N,121.18605431°E		78.72%
(c)-5	31.07380504°N,121.18609173°E	31.07380995°N,121.18609446°E		71.83%
(d)-1	31.07381965°N,121.18616341°E	31.07382170°N,121.18616819°E		
(d)-2	31.07378933°N,121.18611080°E	31.07379395°N,121.18611569°E	0.62± 0.1m	65.91%
(d)-3	31.07376563°N,121.18616652°E	31.07376815°N,121.18617205°E		90.59%
(d)-4	31.07373979°N,121.18610677°E	31.07374529°N,121.18610951°E		67.50%
(d)-5	31.07374750°N,121.18627709°E	31.07374921°N,121.18628438°E		29.29%
(e)-1	31.07401171°N,121.18602609°E	31.07401405°N,121.18603193°E		
(e)-2	31.07392183°N,121.18602088°E	31.07392261°N,121.18602922°E	0.70± 0.08m	71.17%
(e)-3	31.07396429°N,121.18608704°E	31.07396768°N,121.18609214°E		79.53%
(e)-4	31.07386617°N,121.18597500°E	31.07386808°N,121.18598251°E		67.10%
(e)-5	31.07386952°N,121.18611962°E	31.07387298°N,121.18612688°E		54.36%
		Average errors		0.65m

As shown in Table 3, the average ground distance between the UHI and the corrected location have been up to 2.92 meters, but the error of single UHI is reduced to about 1 meter after the proposed method, shown as Table 4. Due to each of the UHI with only one geographic coordinate, these errors are cumulated when we stitch a large number of hyperspectral images, affecting the accuracy of individual tree analysis greatly.

2) THE SPATIAL COORDINATES CORRECTION OF INDIVIDUAL TREE

Hyperspectral images were downloaded from the inner system on Gaia Skymini2. Lenses correction, reflection correction and atmospheric correction were applied using the software provided by the manufacture company. Images are stitched in PHOTOMOD (Racurs, Russia) in following instructions: (1) import camera settings and exterior elements of images; (2) execute aerial triangulation by bundle adjustment with control points if exist; (3) calculate and generate digital elevation model (DEM) of the region; (4) orthorectify images with the DEM generated in (3) and output the whole

region orthographic image. We got both UHI and UVI of the whole region through this procedure.

According to the proposed method, the UHI coordinates are updated to obtain more accurate coordinates of individual tree, and the geographic coordinates of UHI are corrected. Five representative spatial coordinates correction examples of individual tree are shown in Fig 9.

Each hyperspectral image pixel vector is acquired by hundreds of spectral bands and provides very valuable spectral information in species classification and identification. Common machine learning-based classic classifiers, such as k-nearest neighbors [47], logistic regression, support vector machines (SVM) [48] and extreme learning machine (ELM) [49], are employed in hyperspectral classification to achieve satisfactory performance. In this paper, we mainly use manual classification to avoid accumulative errors in the process of coordinate superposition (see Fig 9).

There are two types of errors: (1) the errors between location in the UHI and the actual location of individual tree are shown in Table 5 and (2) the errors between the corrected location and the actual location of individual tree are shown

in Table 6. The locations in UHI are obtained by the stitched hyperspectral image via the software PHOTOMOD, which have cumulative error. As shown in Table 5, the average ground distance between the locations in UHI and the actual location have been up to 2.35 meters, and we can effectively increase the correction ratio to 62.97% by the proposed method. It is found that our method maintains accurate alignments in all experiments, which means that our proposed method can successfully handle the small UAV-based multi-sensor images spatial coordinates correction problem in most time. This offers great possibilities for monitoring species diversity.

V. CONCLUSION

In this paper, we have presented a novel spatial coordinates correction approach by registering low-altitude UAV visible light and hyperspectral images to reduce the positional errors of hyperspectral images captured from a small UAV and provided a precise forest dynamics monitoring. The proposed method first employs visible images and ground control points to stitch a geographic coordinate system as our groundtruth. Hyperspectral images are then registered onto the stitched visible light image via a novel iterative image registration method. Finally, spatial coordinates of the registered hyperspectral images are updated by using the aforementioned groundtruth. Extensive experiments regarding image registration and spatial coordinates correction demonstrate the favorable performance of our method. Our method shows best registration performances against four state-of-the-art registration methods, and effectively increases the correction ratio of individual tree to 62.97%. The fully realization of the dynamic forests monitoring will require automatic monitoring algorithms between the registered images to identify far more new species.

APPENDIX A

THE SOLUTION OF THE OPTIMAL PARAMETER ESTIMATION

Then Equation 4 can be rewritten by ignoring the derivative-redundant terms as

$$\begin{aligned} \mathcal{Q}(\kappa, \Psi, \sigma^2) = & \frac{1}{2\sigma^2} \sum_{j=1}^N \sum_{i=1}^N p_{ij} \|V_{(j,\cdot)} - \mathcal{T}(H, \Psi)_{(i,\cdot)}\|^2 \\ & - N_p \log\left(\frac{\sigma^2 \kappa}{1 - \kappa}\right) - N \log \kappa + \frac{\alpha}{2} \text{tr}(\Psi^T \Gamma \Psi), \end{aligned} \quad (15)$$

where $N_p = \sum_{i=1}^N \sum_{j=1}^N p_{ij}$.

The optimal solution of the transformation is the extrema of the Equation 15. We solve the transformation parameter estimation (M-step) in the matrix form of \mathcal{Q}

$$\begin{aligned} \mathcal{Q} = & \frac{1}{2\sigma^2} \{ \text{tr}(\tilde{V}^T d(P1)\tilde{V}) - 2\text{tr}(H^T P\tilde{H}) \\ & - 2\text{tr}(\Psi^T \Gamma P\tilde{V}) + \text{tr}(H^T d(P^T 1)H) \\ & + 2\text{tr}(\Psi^T \Gamma d(P^T 1)H) + \text{tr}(\Psi^T \Gamma d(P^T 1)\Gamma\Psi) \} \\ & + N_p \log\left(\frac{\sigma^2 \kappa}{1 - \kappa}\right) - N \log \kappa + \frac{\alpha}{2} \text{tr}(\Psi^T \Gamma \Psi), \end{aligned} \quad (16)$$

where $d(\cdot)$ denotes the diagonal matrix formed from a vector. $\mathbf{1}$ is a column vector of filled with ones.

In final, taking the partial derivative of \mathcal{Q} with respect to the parameter Ψ , we obtain

$$\Psi = [\Gamma + \alpha\sigma^2 d(P1)^{-1}]^{-1} [d(P1)^{-1} P V - H] \quad (17)$$

ACKNOWLEDGMENT

(Rui Yu and Minghao Lyu contributed equally to this work.) The authors would like to thank D. G. Lowe, H. Bay, A. Myronenko, Z. Liu, and J. Ma for providing their implementation source codes. These greatly facilitated the comparison experiments. Finally, they would also like to thank the editor and the reviewers for their valuable comments and suggestions.

REFERENCES

- [1] A. Gil-Tena, S. Saura, and L. Brotons, "Effects of forest composition and structure on bird species richness in a Mediterranean context: Implications for forest ecosystem management," *Forest Ecol. Manage.*, vol. 242, nos. 2–3, pp. 470–476, Apr. 2007.
- [2] J. F. Dwyer, E. G. McPherson, H. W. Schroeder, and A. Rowntree, "Assessing the benefits and costs of the urban forest," *J. Arboriculture*, vol. 18, p. 227, Sep. 1992.
- [3] R. Pu, S. Landry, and Q. Yu, "Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery," *Int. J. Remote Sens.*, vol. 32, no. 12, pp. 3285–3308, Jun. 2011.
- [4] R. Pu and S. Landry, "A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species," *Remote Sens. Environ.*, vol. 124, pp. 516–533, Sep. 2012.
- [5] X. Li and G. Shao, "Object-based urban vegetation mapping with high-resolution aerial photography as a single data source," *Int. J. Remote Sens.*, vol. 34, no. 3, pp. 771–789, Feb. 2013.
- [6] L. T. Waser, K. Jutte, M. Kuchler, and T. Stampfer, "Evaluating the potential of WorldView-2 data to classify tree species and different levels of ash mortality," *Remote Sens.*, vol. 6, no. 5, pp. 4515–4545, 2014.
- [7] M. Kamal, S. Phinn, and K. Johansen, "Object-based approach for multi-scale mangrove composition mapping using multi-resolution image datasets," *Remote Sens.*, vol. 7, no. 4, pp. 4753–4783, Apr. 2015.
- [8] J. Arroyo-Mora, M. Kalacska, D. Inamdar, R. Soffer, O. Lucanus, J. Gorman, T. Naprstek, E. Schaaf, G. Ifimov, K. Elmer, and G. Leblanc, "Implementation of a UAV-hyperspectral pushbroom imager for ecological monitoring," *Drones*, vol. 3, no. 1, p. 12, Jan. 2019.
- [9] S. Jay, F. Baret, D. Dutartre, G. Malatesta, S. Héno, A. Comar, M. Weiss, and F. Maupas, "Exploiting the centimeter resolution of UAV multispectral imagery to improve remote-sensing estimates of canopy structure and biochemistry in sugar beet crops," *Remote Sens. Environ.*, vol. 231, Sep. 2019, Art. no. 110898.
- [10] L. Tang and G. F. Shao, "Drone remote sensing for forestry research and practices," *J. Forestry Res.*, vol. 26, pp. 791–797, Dec. 2015.
- [11] S. Wang, L. Liu, L. Qu, C. Yu, Y. Sun, F. Gao, and J. Dong, "Accurate Ulva prolifera regions extraction of UAV images with superpixel and CNNs for ocean environment monitoring," *Neurocomputing*, vol. 348, pp. 158–168, Jul. 2019.
- [12] O. Hassaan, A. K. Nasir, H. Roth, and M. F. Khan, "Precision forestry: Trees counting in urban areas using visible imagery based on an unmanned aerial vehicle," *IFAC-PapersOnLine*, vol. 49, no. 16, pp. 16–21, 2016.
- [13] S. Zhang, K. Yang, Y. Yang, and Y. Luo, "Nonrigid image registration for low-altitude UAV images with large viewpoint changes," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 592–596, Apr. 2018.
- [14] F. Song, T. Dan, R. Yu, K. Yang, Y. Yang, W. Chen, X. Gao, and S.-H. Ong, "Small UAV-based multi-temporal change detection for monitoring cultivated land cover changes in mountainous terrain," *Remote Sens. Lett.*, vol. 10, no. 6, pp. 573–582, Jun. 2019.
- [15] F. Song, Z. Yang, X. Gao, T. Dan, Y. Yang, W. Zhao, and R. Yu, "Multi-scale feature based land cover change detection in mountainous terrain using multi-temporal and multi-sensor remote sensing images," *IEEE Access*, vol. 6, pp. 77494–77508, 2018.

- [16] M. Dalponte, H. O. Ørka, L. T. Ene, T. Gobakken, and E. Næsset, "Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data," *Remote Sens. Environ.*, vol. 140, pp. 306–317, Jan. 2014.
- [17] R. Moreno, F. Corona, A. Lendasse, M. Graña, and L. S. Galvão, "Extreme learning machines for soybean classification in remote sensing hyperspectral images," *Neurocomputing*, vol. 128, pp. 207–216, Mar. 2014.
- [18] X. Yu, J. Hyypää, P. Litkey, H. Kaartinen, M. Vastaranta, and M. Holopainen, "Single-sensor solution to tree species classification using multispectral airborne laser scanning," *Remote Sens.*, vol. 9, no. 2, p. 108, Jan. 2017.
- [19] L. Wei, M. Yu, Y. Liang, Z. Yuan, C. Huang, R. Li, and Y. Yu, "Precise crop classification using spectral-spatial-location fusion based on conditional random fields for UAV-borne hyperspectral remote sensing imagery," *Remote Sens.*, vol. 11, no. 17, p. 2011, Aug. 2019.
- [20] A. Khaliq, L. Comba, A. Biglia, D. R. Aimonino, M. Chiaberge, and P. Gay, "Comparison of satellite and UAV-based multispectral imagery for vineyard variability assessment," *Remote Sens.*, vol. 11, no. 4, p. 436, Feb. 2019.
- [21] H. Rizeei and B. Pradhan, "Urban mapping accuracy enhancement in high-rise built-up areas deployed by 3D-orthorectification correction from WorldView-3 and LiDAR imageries," *Remote Sens.*, vol. 11, no. 6, p. 692, Mar. 2019.
- [22] G. T. Miyoshi, N. N. Imai, A. M. G. Tommaselli, and E. Honkavaara, "Comparison of pixel and region-based approaches for tree species mapping in Atlantic forest using hyperspectral images acquired by UAV," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vols. XLII-2/W13, pp. 1875–1880, Jun. 2019.
- [23] W. M. Xu, J. Y. Wang, R. Shu, Z. P. He, and K. M. Fang, "Theoretical analysis of geometric rectification accuracy for linear pushbroom imaging spectrometer," (in Chinese), *J. Infr. Millim. Waves*, vol. 25, no. 2, pp. 109–112, 2006.
- [24] G. Yang, F. Xie, C. Y. Liu, H. L. Shao, J. Y. Lin, J. Y. Wang, and Y. Guo, "Self-calibration method for eccentric angle in hyperspectral and LiDAR integrated airborne system," *Proc. SPIE*, vol. 10255, Mar. 2017, Art. no. 1025520.
- [25] *The Store Homepage of the DJI Phantom 4 Pro*. Accessed: Jan. 21, 2020. [Online]. Available: <https://www.dji.com/cn/phantom-4-pro>
- [26] *The Store Homepage of the DJI M600 Pro*. Accessed: Jan. 21, 2020. [Online]. Available: <https://www.dji.com/cn/matrice600-pro/info#specs>
- [27] *The Store Homepage of the Gaiasky Mini Pushbroom Hyperspectral Imaging System*. Accessed: Jan. 21, 2020. [Online]. Available: <http://www.dualix.com.cn/Goods/desc/id/123/aid/954>
- [28] C. Aguilera, F. Barrera, F. Lumbreras, A. D. Sappa, and R. Toledo, "Multispectral image feature points," *Sensors*, vol. 12, no. 9, pp. 12661–12672, 2012.
- [29] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–22, 1977.
- [30] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, pp. 325–340, Dec. 1987.
- [31] A. L. Yuille and N. M. Grzywacz, "A mathematical analysis of the motion coherence theory," *Int. J. Comput. Vis.*, vol. 3, pp. 155–175, Jun. 1989.
- [32] F. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-11, no. 6, pp. 567–585, Jun. 1989.
- [33] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
- [34] T. Sikora, "The MPEG-7 visual standard for content description—An overview," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 696–702, Jun. 2001.
- [35] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.
- [36] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognit.*, vol. 46, no. 12, pp. 3519–3532, Dec. 2013.
- [37] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.
- [38] Y. Yang, S. H. Ong, and K. W. C. Foong, "A robust global and local mixture distance based non-rigid point set registration," *Pattern Recognit.*, vol. 48, no. 1, pp. 156–173, Jan. 2015.
- [39] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, Jan. 2016.
- [40] S. Zhang, Y. Yang, K. Yang, Y. Luo, and S. Ong, "Point set registration with global-local correspondence and transformation estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2669–2677.
- [41] K. Yang, A. Pan, Y. Yang, S. Zhang, S. Ong, and H. Tang, "Remote sensing image registration using multiple image features," *Remote Sens.*, vol. 9, no. 6, p. 581, Jun. 2017.
- [42] Z. Wei, Y. Han, M. Li, K. Yang, Y. Yang, Y. Luo, and S.-H. Ong, "A small UAV based multi-temporal image registration for dynamic agricultural terrace monitoring," *Remote Sens.*, vol. 9, no. 9, p. 904, Aug. 2017.
- [43] S. Zhang, K. Yang, Y. Yang, Y. Luo, and Z. Wei, "Non-rigid point set registration using dual-feature finite mixture model and global-local structural preservation," *Pattern Recognit.*, vol. 80, pp. 183–195, Aug. 2018.
- [44] Z. Yang, Y. Yang, K. Yang, and Z.-Q. Wei, "Non-rigid image registration with dynamic gaussian component density and space curvature preservation," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2584–2598, May 2019.
- [45] Z. Yang, T. Dan, and Y. Yang, "Multi-temporal remote sensing image registration using deep convolutional features," *IEEE Access*, vol. 6, pp. 38544–38555, 2018.
- [46] D. Bi, R. Yu, M. Li, Y. Yang, K. Yang, and S. H. Ong, "Multiple image features-based retinal image registration using global and local geometric structure constraints," *IEEE Access*, vol. 7, pp. 133017–133029, 2019.
- [47] E. Blanzieri and F. Melgani, "Nearest neighbor classification of remote sensing images with the maximal margin principle," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1804–1811, Jun. 2008.
- [48] J. Liu, Z. Wu, Z. Wei, L. Xiao, and L. Sun, "Spatial-spectral kernel sparse representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 6, pp. 2462–2471, Dec. 2013.
- [49] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.



RUI YU is currently pursuing the M.S. degree with the School of Information Science and Technology, Yunnan Normal University. Her current research interests include remote sensing image registration, change detection, and medical imaging.



MINGHAO LYU is currently pursuing the bachelor's degree with the Geographic Information Science in Forestry College, Northeast Forestry University. He is also a Visiting Student with the Engineering Research Center of GIS Technology in Western China, Ministry of Education, Yunnan Normal University, under the supervision of Prof. Y. Yang. His research interest includes population structure and dynamics of vegetation in natural forest using remote sensing technology.



JIAHUI LU is currently pursuing the M.S. degree with the School of Environmental and Ecological Science, East China Normal University. His research interests include remote sensing of vegetation and deep learning application in ecology.



YANG YANG received the master's degree from Waseda University, Japan, in 2007, and the Ph.D. degree from the National University of Singapore, Singapore, in 2013. He is currently a Professor with the School of Information Science and Technology, Yunnan Normal University. His research interests include computer vision, remote sensing, geography information systems, and medical imaging.



FEI LI received the master's degree from Yunnan University, in 2010. He is currently an Associate Professor with the Department of Scientific Research, Yunnan Normal University. His research interests include mathematical modeling and scientific computing, and image processing.

• • •



GUOCHUN SHEN received the Ph.D. degree from Zhejiang University, China, in 2010. He is currently an Associate Professor with the School of Ecological and Environmental Sciences, East China Normal University. At present, he is conducting joint research project with the Engineering Research Center of GIS Technology in Western China, Yunnan Normal University. His research focuses on how and why species, especially subtropical forest trees, coexist in heterogeneous environments, and how can we accurately and robustly detect and estimate effects of abiotic habitat, dispersal, individual interaction on community structures from a spatially explicit perspective.