remote sensing

MDPI

Article

A Small UAV Based Multi-Temporal Image Registration for Dynamic Agricultural Terrace Monitoring

Ziquan Wei [1,2,†], Yifeng Han [1,2,†], Mengya Li [1,2,†], Kun Yang [1,3,*], Yang Yang [1,2,3,*], Yi Luo [1,3,*] and Sim-Heng Ong [4]

[1] School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China; wzq10101@gmail.com (Z.W.); wlpspx@ynnu.edu.cn (Y.H.); glorialmy@outlook.com (M.L.)

[2] Laboratory of Pattern Recognition and Artificial Intelligence, Yunnan Normal University, Kunming 650500, China

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

[4] Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576, Singapore; eleongsh@nus.edu.sg

* Correspondence: kmdcynu@163.com (K.Y.); yyang_ynu@163.com (Y.Y.); luoyi861030@163.com (Y.L.)

† These authors contributed equally to this work.

Received: 15 July 2017; Accepted: 21 August 2017; Published: 31 August 2017

Abstract: Terraces are the major land-use type of agriculture and support the main agricultural production in southeast and southwest China. However, due to smallholder farming, complex terrains, natural disasters and illegal land occupations, a light-weight and low cost dynamic monitoring of agricultural terraces has become a serious concern for smallholder production systems in the above area. In this work, we propose a small unmanned aerial vehicle (UAV) based multi-temporal image registration method that plays an important role in transforming multi-temporal images into one coordinate system and determines the effectiveness of the subsequent change detection for dynamic agricultural terrace monitoring. The proposed method consists of four steps: (i) guided image filtering based agricultural terrace image preprocessing, (ii) texture and geometric structure features extraction and combination, (iii) multi-feature guided point set registration, and (iv) feature points based image registration. We evaluated the performance of the proposed method by 20 pairs of aerial images captured from Longji and Yunhe terraces, China using a small UAV (the DJI Phantom 4 Pro), and also compared against four state-of-the-art methods where our method shows the best alignments in most cases.

Keywords: agricultural terrace; dynamic monitoring; multi-temporal; image registration; small UAVs

1. Introduction

China has more than 70% of its land made up of mountains and hills, while the main agricultural terraces are located in southwest and southeast China. Therefore, agricultural terraces have become the most important agricultural land-use type and support the main agricultural production in these areas. Furthermore, it also plays an important role in reducing flood runoff, changing terrain slope and maintaining water, soil and conservation fertilizer.

However, due to Chinese agricultural policy and regional characteristics, most agricultural terraces in southwest and southeast China are farmed by smallholders, and have small sizes, scattered distributions, complex terrains and other characteristics. Such issues have increased difficulty in land use management and planting management of local governments, as well as regular cropping and planting area monitoring of smallholder farmers. Meanwhile, illegal land occupations, natural disasters, water and wind erosion are also causing a drastic decrease in agricultural terraces, and these further aggravate the soil erosion and endanger the smallholder production. Consequently, a light-weight and low cost dynamic monitoring of agricultural terraces has become a serious concern for smallholder production systems in China.

Agricultural terrace monitoring includes two major aspects: (i) crop monitoring and (ii) planting area monitoring. In this work, we mainly focus on the planting area monitoring, which generally depends on aerial remote sensing and image processing techniques. In parallel to the technological advances of aerial remote sensing, achievements in image processing have promoted the development of sophisticated algorithms using aerial images, such as the landform classification, which is one of the effective ways for dynamic monitoring of land-use changes. Traditional landform classification methods can be divided into non-supervised classification methods and supervised classification methods [1]. Basically, non-supervised methods consist of manual classification methods and automated classification methods [2]. Manual classification methods are relatively time-consuming and the results depend on subjective decisions of the interpreter and are, therefore, neither transparent nor reproducible [3,4]. Automated classification methods [5,6] make use of the unsupervised nature and automation of the change analysis process. However, they are unfavored by difficulties in identifying and labeling change trajectories [7], and the lack of information on calibration of the ground [7]. Artificial neural network (ANN) constitutes a key component of supervised classification methods [8,9]. It is a non-parametric method that is capable of estimating the properties of data based on the training samples. However, ANN suffers from the long training time, the sensitivity to the amount of training data used, as well as the applicability of ANN functions in the common image processing softwares [7].

Subsequent to the landform classification, classified images can be adopted to monitor land use changes. Over the last few years, in order to analyze agricultural terraces, different and new technologies have been used. Satellite data of high spatial resolution and advanced image processing techniques, have opened up a new insight for mapping landscape features, such as terraces. This has given the opportunity for quantitative assessment of farming practices as an indicator in water pollution risk assessment [10], soil erosion risk assessment [11] and landslide boundary monitoring [11–14]. In addition, airborne LiDAR (light detection and ranging), which has been developed to collect and subsequently characterize vertically distributed attributes [15], and the derived digital elevation model (DEM) [16–20] or digital terrain model (DTM) [12,14] are becoming standard practices in spatial related areas. Recently, the use of unmanned aerial vehicle (UAV) for civil applications has emerged as an attractive and flexible option for the monitoring of various aspects of agriculture and environment [21]. For example, Diaz-Varela et al. [21] proposed an automatic identification of agricultural terraces through object-oriented analysis of high resolution digital surface models and multi-spectral images obtained from UAVs. Deffontaines et al. [22] monitored the active inter-seismic shallow deformation of the Pingting terraces by using UAV high resolution topographic data combined with InSAR time series. Yang et al. [23] proposed a multi-viewpoint remote sensing image registration method that provided an accurate mapping between different viewpoint images for ground change detections.

Compared with satellite and other aerial remote sensing, using small UAVs for agricultural terrace monitoring has a strong mobility, high efficiency, low cost and other advantages. However, the following issues still exist: (i) Due to the payload capacity, small UAVs usually can only carry a light-weight visible light camera, such as CCD or CMOS cameras that limit available image information while increasing difficulty in monitoring algorithms compared with using multi-spectral imaging. (ii) When collecting multi-temporal images for the same location (e.g., a planting area in terraces), the imaging perspective of small UAVs is often easily affected by wind speed/direction, complex terrain, battery capacity (e.g., flying distance), aircraft posture (pitch, roll, yaw), flying height and other human factors. These factors cause the captured scenes (i.e., the same location in a pair of multi-temporal images) to not be in the same coordinate system, while image geometric distortions, low image overlapping, brightness changes and color changes may also be produced in such multi-temporal images.

The above issues have led to the fact that multi-temporal images of the same scene captured by small UAVs may not be directly used to detect changes for dynamic agricultural terrace monitoring, and a reliable multi-temporal image registration, which can transform the images into one coordinate system, is necessary in order to be able to subsequently compare or integrate the data obtained from the multi-temporal images.

In this work, we focus on planting areas of agricultural terraces, and present a small UAV based multi-temporal image registration method for dynamic agricultural terrace monitoring. The major contributions of the proposed method includes: (i) the guided image filtering for agricultural terrace image preprocessing is first designed to enhance terrace ridges in multi-temporal images, (ii) the multi-feature descriptor is then applied to combine the texture feature and the geometric structure feature of terrace images for improving the description of feature points and rejecting outliers, (iii) the multi-feature guided model provides an accurate guiding for feature point set registration, and (iv) the feature points based image registration finally registers the terrace images accurately.

2. Methodology

The proposed small UAV based multi-temporal image registration method has four major sequential processes: (i) image preprocessing; (ii) feature extraction and combination; (iii) feature point set registration; and (iv) image registration. In this section, we first introduce the proposed method followed by analyzing the computational complexity and discussing the implementation details.

2.1. Guided Image Filtering Based Agricultural Terrace Image Preprocessing

Given a gray agricultural terrace image **I** with $x \times y$ pixels and intensity **I** = 0.3R + 0.59G + 0.11B from the colorized image captured by a small UAV. We first define a preprocessing method to strengthen the identifiability of terrace ridges in multi-temporal images, and then extract salient features of terrace images along the enhanced ridges in the second step. The main reason is that accuracy and validity of an image registration are not only controlled by the performance of feature point set registration, but also determined by the number and the distribution density of feature points, because of the image transformation constructed by abundant feature points.

In this work, the preprocessing method improves the contrast ratio between terraces and their ridges using the Guided Image Filtering (GIF) [24]. In order to extract large and quality feature points from terrace ridges, we first adopt the GIF, which has the edge-preserving smoothing and the gradient preserving, to preprocess input images. The GIF employs a guidance image to construct a spatially variant kernel and is also related to the matting Laplacian matrix [25].

Firstly, a linear translation-variant guided filtering process in a square window π_k centered at a pixel k is defined by:

$$\Lambda_i = \alpha_k \mathbf{I}_i + \beta_k, \forall i \in \pi_k, \tag{1}$$

where i is a pixel index, Λ is the output image, $\alpha_k = \frac{\frac{1}{|\pi|}\sum_{i\in\pi_k}\mathbf{I}_i g_i - \gamma_k \overline{g}_k}{\delta_k^2+\varepsilon}$ and $\beta_k = \overline{g}_k - \alpha_k \gamma_k$ are two parameters of the minimal cost function $\underset{\alpha_k,\beta_k}{\arg\min}\sum_{i\in\pi_k}((\alpha_k \mathbf{I}_i + \beta_k - g_i)^2 + \varepsilon\alpha_k^2)$. Here, g is the input image that is identical to the guidance image **I**, γ_k and δ_k^2 are the mean and the variance of **I** in π_k, $|\pi|$ is the number of pixels in π_k, ε is a regularization parameter preventing α_k from being too large, and $\overline{g}_k = \frac{1}{|\pi|}\sum_{i\in\pi_k} g_i$ is the mean of g in π_k.

Secondly, we apply the linear model to all local windows in the entire image:

$$\begin{aligned}\Lambda_i &= \frac{1}{|\pi|}\sum_{k:i\in\pi_k}(\alpha_k \mathbf{I}_i + \beta_k) \\ &= \overline{\alpha}_i \mathbf{I}_i + \overline{\beta}_i,\end{aligned} \tag{2}$$

where $\overline{\alpha}_i = \frac{1}{|\pi|} \sum_{k \in \pi_k} \alpha_k$ and $\overline{\beta}_i = \frac{1}{|\pi|} \sum_{k \in \pi_k} \beta_k$. An example of agricultural terrace image enhancement (250 × 150 pixels) using GIF is given in Figure 1.

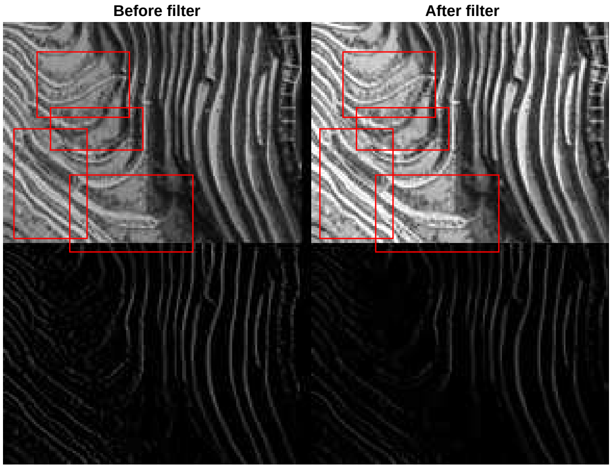


Figure 1. An example of guided image filtering. The images before and after filtering are shown in the first row, and the enhanced ridges are marked by the red windows. The image gradients before and after filtering are shown in the last row, exhibiting the gradient preserving by the GIF.

An agricultural terrace is constructed by cropland and terrace ridges. Ridges can be considered as a salient feature that provides the features of geometric contours and surface textures for feature based terrace image registration. However, the flat color of terrace ridge is similar with croplands and detecting cropland and terrace ridges becomes difficult when crops are growing in the early stage. Thus, the extracted feature points along ridges are more helpful than points distributed on the cropland. Mathematically, the exponential function can change the density of the data distribution; therefore, we expand the gray value distribution of the common terrace via a natural exponential function formed as:

$$\mathbf{I}_i^{new} = \begin{cases} 1, & \text{if } exp(-\Lambda_i) \geq 0.7, \\ exp(-\Lambda_i), & \text{otherwise.} \end{cases} . \tag{3}$$

Finally, we can obtain the preprocessed gray agricultural terrace image $\mathbf{I}^{new}$ with a high contrast ratio, smoothing edges, and prominent ridges. The preprocessing step gives an opportunity to extract quality features from these preprocessed terrace images.

2.2. Features Extraction and Combination

Feature points are selected by using the good-feature-to-track criterion [26], similar to the Harris detector, based on the second moment matrix [27]. The selection specifically maximizes the quality of tracking, and is therefore optimized by construction, as opposed to more ad hoc measures of texturedness. The selected feature point set $\mathbf{P} = \{\mathbf{p}_t\}_{t=1}^{T}$ belongs with the geometric coordinate of a input agricultural terrace image pixel, where $\{\mathbf{p}_t\}_{t=1}^{T} \in \mathbb{Z}^+$. There is an example of feature points extraction from a preprocessed agricultural terrace image (500 × 300 pixels) as shown in Figure 2.

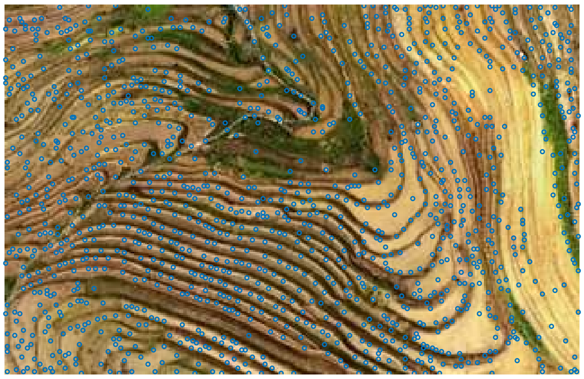


Figure 2. An example of feature points extraction. There are 1270 feature points extracted and denoted by blue circles.

2.2.1. Local Texture Feature Descriptor

A local texture (LT) feature descriptor is designed to describe the texture features around each feature point according to the dominant rotated local binary patterns (DRLBP) proposed by Mehta [28]. Given a gray image **I** with $x \times y$ pixels. The DRLBP operates in a local circular region by taking the difference of the central pixel with respect to its neighbors. It is defined as:

$$DRLBP^{\mathbf{I}}_{R,L}(x,y) = \sum_{l=0}^{L-1} m(i(x,y), i(a_l(x), b_l(y))) \cdot 2^{mod(l-D,L)}, \tag{4}$$

where

$$m(i(x,y), i(a_l(x), b_l(y))) = \begin{cases} 1, & i(a_l(x), b_l(y)) \geq i(x,y), \\ 0, & i(a_l(x), b_l(y)) < i(x,y). \end{cases},$$

$i(x,\ y)$ and $i(a_l(x),\ b_l(y))$ are the gray values of central pixel and its neighbor in image **I**, respectively, $(x,\ y)$ and $(a_l(x),\ b_l(y))$ are the geometric coordinate of the central pixel and its lth neighbor. Note that $a_l(x) = x + Rcos(2\pi l/L)$ and $b_l(y) = y - Rsin(2\pi l/L)$, R is the radius of the circular neighborhood and L is the number of the neighbors. The *mod* indicates the modulus operator, and $D = \underset{l \in \{0,1,\ldots,L-1\}}{\arg\max} |i(a_l(x),\ b_l(y)) - i(x,y)|$. In this paper, the $\{DRLBP^{\mathbf{I}}_{R,L}(x,\ y)\}_{x=0,y=0}^{X-1,Y-1}$ are held in binary form. The DRLBP descriptor $DRLBP^{\mathbf{I}}_{R,N}$ of **I** is denoted by the DRLBP histogram $\mathbf{H}_{R,L}(DRLBP^{\mathbf{I}}_{R,L}) \in \mathbb{R}^{2^L}$. The *mod* operator circularly shifts the weights with respect to the dominant direction because of the weight term $2^{mod(l-D,L)}$ depends on D in the above definition. Therefore, the DRLBP is a rotation invariance and computationally efficient texture descriptor.

Before giving the LT for each feature point, the image is first weighted for each feature point based on its geometric coordinates, the weighting matrix for each point $\mathbf{p}_t$ is defined as:

$$\epsilon^t_{xy} = exp(-\frac{||\mathbf{I}_{xy} - \mathbf{p}_t||^2}{2\tau^2}), \tag{5}$$

where $\mathbf{I}_{xy}$ is the geometric coordinate of the pixel $\mathbf{I}(x,y)$, $\mathbf{p}_t$ is the tth point of $\mathbf{P}$, and τ is a parameter that controls the window size of LT. ϵ^t is a $x \times y$ weighting matrix of the tth feature point, and the tth weighting gray image is obtained by:

$$\mathbf{I}^t(x,y) = \epsilon^t_{xy} \times \mathbf{I}(x,y), \tag{6}$$

where the weighting gray image $\mathbf{I}^t$ has the same size with the source gray image. Let $\mathbf{I}^t(x,y) = 0$ when $\mathbf{I}^t(x,y) \leq 10^{-4}$. Figure 3 shows an example of weighting gray image.

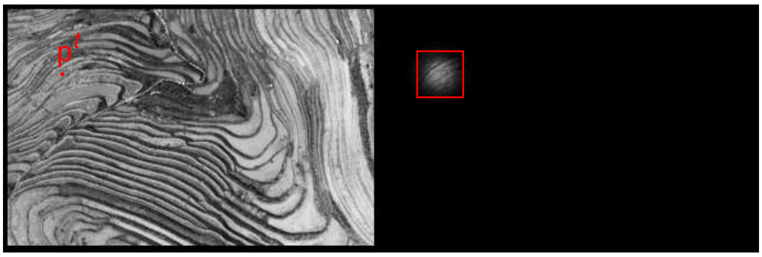


Figure 3. An example of weighting gray image that is based on a feature point $\mathbf{p}^t$. The left one is the original terrace gray image, and the right is the weighting gray image. The effect of the weight operator is shown in a red window, and the size of the window is controlled by the parameter τ^2.

We define the LT of the tth feature point via DRLBP in the weighting gray image:

$$\mathcal{LT}(\mathbf{p}_t)_{R,L} = \mathbf{H}_{R,L}(DRLBP^{\mathbf{I}^t}_{R,L}), \tag{7}$$

where $\mathcal{LT}(\mathbf{P})_{R,L}$ is a LT feature set with T vectors both size of 1×2^L.

2.2.2. Local Geometric Structure Feature Descriptor

A local geometric structure (LGS) feature is designed for each feature point in $\mathbf{P}$ by a local vector weighting method defined by:

$$\mathcal{LGS}(\mathbf{p}_t) = \sum_{k=1}^{K} \eta_{tk} \overrightarrow{\mathbf{p}_t \mathbf{p}_{t_k}}, \tag{8}$$

where t_k is the index of kth neighbor of $\mathbf{p}_t$, and K is the number of neighbors. The LGS descriptor employs the K neighbors to give a LGS description for each feature point $\mathbf{p}_t$ but ignores outlier points. Hence, the value of K and the weight term η_{tk} play a crucial role for the performance of LGS. We use an outlier score to define the weight term of $\overrightarrow{\mathbf{p}_t \mathbf{p}_{t_k}}$:

$$\eta_{tk} = \frac{1}{2\pi\sigma^2_{\mathcal{LGS}}} exp(-\frac{\Delta^{\mathcal{LT}}_{tk}}{2\pi\sigma^2_{\mathcal{LGS}}}), \tag{9}$$

where $\sigma^2_{\mathcal{LGS}}$ is the variance of $\{\Delta^{\mathcal{LT}}_t\}^T_{t=1}$, and the outlier score $\Delta^{\mathcal{LT}}_{t_k}$ is computed by the LT distance between $\mathbf{p}_{t_k}$ and the point that has the most similar LT feature with $\mathbf{p}_{t_k}$ as:

$$\Delta^{\mathcal{LT}}_{tk} = \min \|\mathcal{LT}(\mathbf{p}_{t_k})_{R,L} - \mathcal{LT}(\mathbf{p}_{t(t \neq t_k)})_{R,L}\|^2. \tag{10}$$

2.2.3. Multi-Feature Descriptor

Different types of feature descriptors have their own advantages and limitations. This motivates us to make the respective advantages of LT and LGS descriptors complementary to each other. The multi-feature (MF) is designed to combine the local texture information and the local geometric structure information for improving the identifiability of each feature point. However, a fixed falseness does no good for guiding point registration throughout the iterations. Thus, the MF descriptor is defined as:

$$\mathcal{MF}(\mathbf{P}) = \mathbf{P} + T_1 \mathcal{LGS}(\mathbf{P}) + T_2 \mathcal{LT}(\mathbf{P})_{R,L}, \tag{11}$$

where T_1 and T_2 are annealing parameters for the LGS and the LT features, respectively. The instantiation of Equation (11) is given in the implementation details section.

2.3. Multi-Feature Guided Point Set Registration Model

For feature based image registration, two sets of feature points are extracted from a pair of multi-temporal agricultural terrace images (i.e., a sensed image and a reference image), respectively. The extracted feature points contain a large number of outliers that limit the performance of current non-rigid point set registration algorithms [29–31]. For this issue, a robust multi-feature guided model is designed—given two point sets $\mathbf{A} = \{\mathbf{a}_n\}_{n=1}^{N}$ (i.e., the source point set) and $\mathbf{B} = \{\mathbf{b}_m\}_{m=1}^{M}$ (i.e., the target point set) which are extracted from the sensed image and the reference image, respectively. The proposed point set registration model is first (i) to estimate correspondences between **A** and **B** by the proposed MF descriptor at each iteration, and then (ii) to update the location of **A** using a non-rigid transformation built by the recovered correspondences. The steps (i) and (ii) are iterated such that the **A** can gradually and continuously approach the target point set **B** , and finally match the exact corresponding points in **B** .

2.3.1. Correspondence Estimation

In the first step, the Gaussian mixture model (GMM) is applied to estimate correspondences by measuring the similarity of the MF between two point sets, and the correspondence estimation problem is considered as a GMM probability density estimation problem. Let the MF of $\mathbf{a}_n$ be the centroid of the nth Gaussian component, and the MF of $\mathbf{b}_m$ be the mth data. The GMM probability density function (PDF) is therefore obtained as:

$$s(\mathcal{MF}(\mathbf{b}_m)) = (1-\omega)\sum_{n=1}^{N}\mathcal{P}_{mn}\phi(\mathcal{MF}(\mathbf{b}_m)|\mathcal{MF}(\mathbf{a}_n)) + \frac{\omega}{M}, \tag{12}$$

where $\phi(\mathcal{MF}(\mathbf{b}_m)|\mathcal{MF}(\mathbf{a}_n)) = \frac{1}{2\pi\sigma^2}exp(-\frac{\|\mathcal{MF}(\mathbf{b}_m)-\mathcal{MF}(\mathbf{a}_n)\|^2}{2\sigma^2})$ with the equal isotropic covariances σ^2 of MF, $\mathcal{P}_{mn} = \frac{1}{N}$ are non-negative equal quantity with $\sum_{n=1}^{N}\mathcal{P}_{mn} = 1$, which are called the priors of GMM. $\frac{\omega}{M}$ is an additional uniform distribution with a weighting parameter ω, $0 < \omega < 1$ for outlier dealing.

Once we have the PDF of GMM that is guided by the similarity of MFs, we can estimate correspondence by the posterior probability of GMM via Bayes' rule:

$$s_{nm}(\mathcal{MF}(\mathbf{a}_n)|\mathcal{MF}(\mathbf{b}_m)) = \frac{exp(-\frac{\|\mathcal{MF}(\mathbf{b}_m)-\mathcal{MF}(\mathbf{a}_n)\|^2}{2\sigma^2})}{\sum_{i=1}^{N}exp(-\frac{\|\mathcal{MF}(\mathbf{b}_m)-\mathcal{MF}(\mathbf{a}_i)\|^2}{2\sigma^2}) + \frac{2\pi\sigma^2\omega N}{M(1-\omega)}}, \tag{13}$$

by which we obtain an one-to-many fuzzy correspondence matrix $\mathbf{S}_{N\times M}$ guided by the similarity of MFs. Meanwhile, the corresponding target point set is obtained by:

$$\overline{\mathbf{B}} = \mathbf{S}\times\mathbf{B}. \tag{14}$$

The proposed correspondence estimation method imitates the process of human practice, which measures the similarities of geometric structure feature and local texture feature. Generally, the process for humans to estimate the corresponding point of the source point a in terrace image consists of two parts: (i) searching for a region in a reference image that has a similar geometrical location and structure compared to the region that surrounds source point a, and (ii) finding a point within this region that has similar color features (LT in this paper) to the source point set. An example is shown in Figure 4, where only 100 feature points are shown for visual convenience. The two feature points a and b both have the similar pattern of LT histogram and the similar geometric feature.

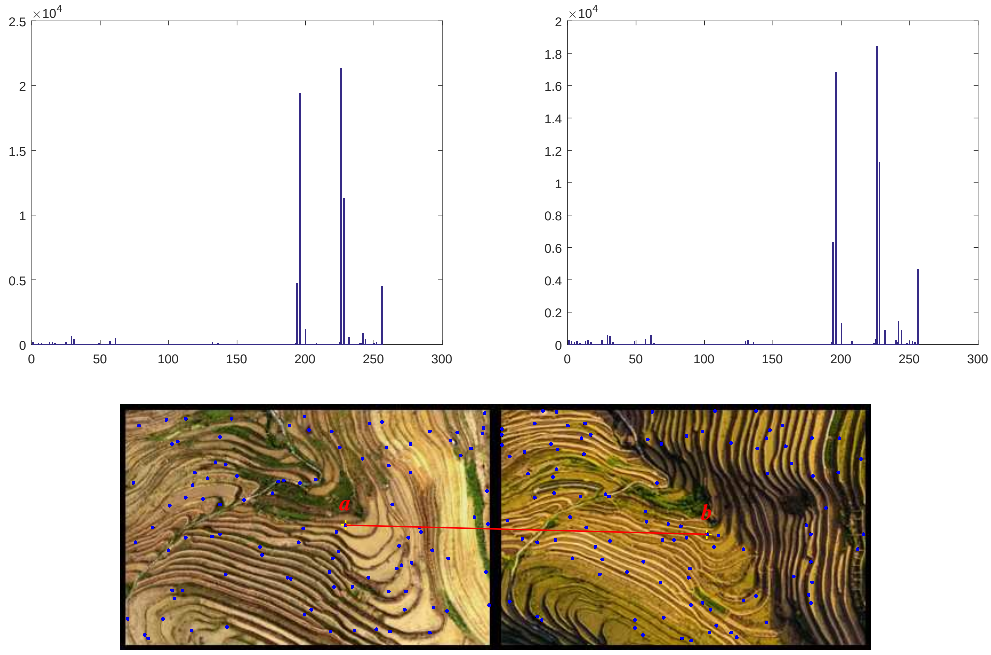


Figure 4. Example on estimating correspondence between two terrace images. In the first row, histograms of the points **a** and **b** are provided, respectively. In the second row, the points **a** and **b** are classified as the corresponding pair, which are connected by the red line.

2.3.2. Transformation Estimation

We model the non-rigid displacement function f by requiring it to lie within a specific functional space, namely a vector-valued reproducing kernel Hilbert space (RKHS) [32,33]. The Gaussian kernel, which is in the form $\mathbf{G}(\mathbf{a}_{n1}, \mathbf{a}_{n2}) = exp(-\frac{1}{2\alpha^2}\|\mathbf{a}_{n1} - \mathbf{a}_{n2}\|^2)$ and of size $N \times N$, is chosen to be the associated kernel for the RKHS, where α is a constant to control the spatial smoothness. The function f can be defined by:

$$f(\mathbf{A}) = \mathbf{A} + \mathbf{G}W. \tag{15}$$

Thus, the transformation estimation boils down to finding a finite parameter matrix W.

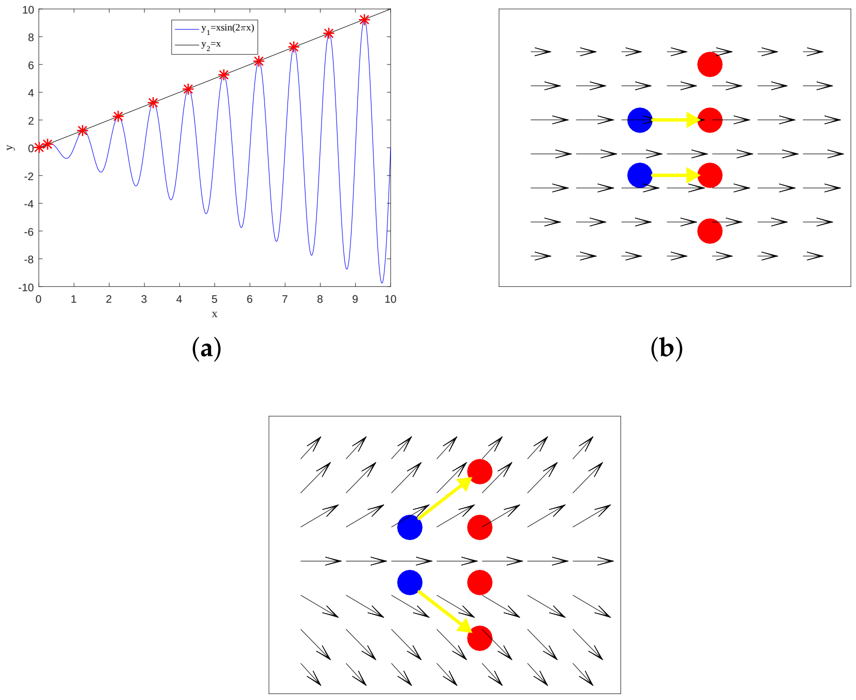
Before a direct parameter estimation, we first illustrate a rule by which a reliable transformation parameter is obtained in the estimation process.

- Regularizing the transformation estimation process.
  The adopted Tikhonov regularization framework [34–37] is one of the most common forms of regularization. It minimizes an energy function in an RKHS $\mathcal{H}$ to regularize a function f, and can be written as:

  $$\mathcal{R}(f) = \|f\|^2_{\mathcal{H}}. \tag{16}$$

  In this paper, the function f is defined in Equation (15).

As shown in Figure 5a, the regularized function (denoted by black line) is more reasonable than its non-regularized counterpart (denoted by blue curve). The transformation of an iterative registration method is such a procedure that slowly displaces the source point set so that the correspondence estimation is easier and more reliable. In other words, regularizing the transformation is necessary to accomplish the iterative registration. Figure 5b,c indicate that the ill-posed problem will exist if the transformation is not regularized. Note that as the number of points increases, the increasing arbitrariness of the transformation will lead to more severe ill-posed problems.



(**a**) (**b**)

(**c**)

Figure 5. Two examples for demonstrating the importance of regularization. (**a**) two different results for a function estimation problem. The black line and blue curve denote the estimated functions with or without being regularized, respectively. In addition, the red asterisks denote 11 data points; (**b**) point set transformation and its velocity field in regularized scenario; (**c**) point set transformation and its velocity field in a non-regularized scenario. In (**b**,**c**), the blue and red points denote the source and target point, respectively.

The multi-feature guided fuzzy correspondence matrix $\mathbf{S}$ contains $N \times M$ probabilities, hence a reliable transformation will produce a larger expectation of probabilities. Therefore, the solution of transformation estimation is detected by maximizing a likelihood function that is formed as $\Pi_{m=1}^{M} s(\mathcal{MF}(\mathbf{b}_m))$, or equivalent to minimizing the negative log-likelihood function, which is formed as:

$$E(W,\sigma^2) = -\sum_{m=1}^{M} log(1-\omega) \sum_{n=1}^{N} \mathcal{P}_{mn}\phi(\mathcal{MF}(\mathbf{b}_m)|\mathcal{MF}(\mathbf{a}_n)) + \frac{\omega}{M}. \tag{17}$$

We use the maximizing expectation (M-step) of the expectation maximization (EM) algorithm [29] to estimate the transformation. The idea of the EM algorithm is first to guess the values of parameters ("old" parameter values) via computing the posterior probability by Equation (13) (E-step), and then to find the "new" parameter values via minimizing the expectation of the complete negative log-likelihood function (M-step), which is formed as:

$$Q(W,\sigma^2) = \frac{1}{2\sigma^2} \sum_{m=1}^{M} \sum_{n=1}^{N} s_{nm} \|\mathcal{MF}(\mathbf{b}_m) - \mathcal{MF}(f(\mathbf{a}_n))\|^2 + N_{\mathbf{S}} log\sigma^2 + \frac{\mu}{2}\mathcal{R}(f), \tag{18}$$

where $N_{\mathbf{S}} = \sum_{m=1}^{M} \sum_{n=1}^{N} s_{nm} \leq M$ (with $M = N_{\mathbf{S}}$ only if $\omega = 0$), $\mathcal{R}$ is the regularization of the transformation, and μ is a weighting parameter controlling the strength of the regularization. Furthermore, with an initialized deterministic annealing parameter σ^2, the parameter W is obtained by $\underset{W}{\arg\min}\, Q$. The mathematical solution is detailed in Section 2.5.

2.4. Feature Points Based Image Registration

Let $\mathbf{I}$ and $\mathbf{I}^t$ be the sensed and reference images, where the source point set $\mathbf{A}$ and target point set $\mathbf{B}$ are extracted from $\mathbf{I}$ and $\mathbf{I}^t$, respectively. $x(\cdot) \times y(\cdot)$ denotes the size of one image. Our goal is to obtain the transformed image $\hat{\mathbf{I}}$.

After the transformed source point set $\hat{\mathbf{A}}$ is obtained, a mapping function can be estimated based on the corresponding set that is constructed by $\mathcal{C} = \{\mathbf{A},\ \hat{\mathbf{A}}\}$, and then the image registration can be realized. There are two types of mapping: (i) forward approach: directly transforming the sensed image $\mathbf{I}$ using the mapping function, and (ii) backward approach: determining the transformed image $\hat{\mathbf{I}}$ from $\mathbf{I}$ using the grid of the reference image $\mathbf{I}^t$ and the inverse of the mapping. Due to the discretization and rounding, (i) is complicated to implement, as it can produce holes and/or overlaps in the output image, and we use the backward approach for image transformation.

We employ the TPS (thin plate spline) transformation model which obtained by:

$$\mathbf{E}^{TPS} = \begin{pmatrix} \mathcal{K} & \Phi \\ \Phi^T & \mathbf{O} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{A} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \tag{19}$$

where the TPS model $\mathbf{E}^{TPS}$ is of size $(N+3) \times 3$, $\mathbf{O}$ is a 3×3 matrix of zeros and Φ is the $N \times 3$ matrix with the nth denoting $(1, \hat{\mathbf{a}}_n)$, and the $N \times N$ TPS kernel $\mathcal{K}_{ij} = \|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|^2 log\|\hat{\mathbf{a}}_i - \hat{\mathbf{a}}_j\|$.

A regular grid $\Theta^t_{Z\times 2} = \{\theta^t_z\}_{z=1}^{Z}{}^T$ is obtained by a pixel-by-pixel indexing process on the reference image $\mathbf{I}^t$, where $Z = X(\mathbf{I}^t) \times Y(\mathbf{I}^t)$. Letting grid Θ^t be the source point set, and $\mathbf{E}^{TPS}$ the TPS transformation model, the transformed grid is obtained by first computing

$$\hat{\Theta}^t_{Z\times 3} = \begin{pmatrix} \overline{\mathcal{K}} & \overline{\Phi} \end{pmatrix} \mathbf{E}, \tag{20}$$

then restoring the dimension of the grid to 2 by $\hat{\Theta}^t \leftarrow \begin{pmatrix} \hat{\Theta}^t_{(\cdot,1)} & \hat{\Theta}^t_{(\cdot,2)} \end{pmatrix}$, where the $Z \times N$ kernel $\overline{\mathcal{K}}_{ij} = \|\theta^t_i - \hat{\mathbf{a}}_j\|^2 \log \|\theta^t_i - \hat{\mathbf{a}}_j\|$, $\overline{\Phi}$ is the $Z \times 3$ matrix with the zth row denotes $(1, \theta^t_z)$ and $\hat{\Theta}^t_{(\cdot,i)}$ denotes the ith column of $\hat{\Theta}^t$. Let Θ be the grid obtained on $\mathbf{I}$, we have

$$\hat{\Theta} = \hat{\Theta}^t \cap \Theta. \tag{21}$$

Finally, the transformed image $\hat{\mathbf{I}}$ is obtained by resampling intensities from the sensed image $\mathbf{I}$ based on $\hat{\Theta}$, setting the rest of pixels to black. Note that the bicubic interpolation is used to improve the smoothness of $\hat{\mathbf{I}}$; to be more precise, the intensities of each pixel in $\hat{\mathbf{I}}$ is determined by summing the weighted neighbor pixel intensities within a 4×4 window.

2.5. Implementation Details

The instantiation of Equation (11) is impeded by LT descriptor $\mathcal{LT}$, which is constructed by a 2^L dimension histogram. Actually, the goal of Equation (11) is to measure the distance of MFs, which means that instantiating

$$\Gamma_{nm} = \|\mathcal{MF}(\mathbf{a}_n) - \mathcal{MF}(\mathbf{b}_m)\|$$

is equivalent to instantiating Equation (11). Analogically, $\Gamma_{nm} = \|(\mathbf{a}_n - \mathbf{b}_m) + (\mathcal{LGS}(\mathbf{a}_n) - \mathcal{LGS}(\mathbf{b}_m)) + (\mathcal{LT}(\mathbf{a}_n) - \mathcal{LT}(\mathbf{b}_m))\|$ has two-dimensional terms and one 2^L dimension term, hence, respectively instantiating

$$\mathcal{G}_{nm}(\mathbf{B}, \mathbf{A}) = (\mathbf{b}_m - \mathbf{a}_n) + (\mathcal{LGS}(\mathbf{b}_m) - \mathcal{LGS}(\mathbf{a}_n)) \tag{22}$$

and

$$\Psi_{nm}(\mathbf{B},\mathbf{A}) = \mathcal{LT}(\mathbf{b}_m) - \mathcal{LT}(\mathbf{a}_n) \tag{23}$$

is equivalent to instantiating Equation (11). The distance of geometric and similarity of LT are denoted by Equations (22) and (23), in which the instantiation of $\mathcal{G}$ can be commonly realized by Euclidean distance computation. Furthermore, in general, a human similarity estimation of LT is to differentiate the pattern of the histogram, as we discussed in Section 2.3.1. Therefore, we instantiate Ψ by firstly normalizing each histogram in $[0, 1]$, and then computing the quadratic distance sum of each dimension in the histogram, which is formed as:

$$\begin{aligned} \Psi_{nm} &= \mathcal{LT}(\mathbf{a}_n) - \mathcal{LT}(\mathbf{b}_m) \\ &= \sum_{i=1}^{2^L} (\mathcal{LT}(\mathbf{a}_n, i) - \mathcal{LT}(\mathbf{b}_m, i))^2, \end{aligned} \tag{24}$$

where $\mathcal{LT}(\mathbf{a}_n, i)$ denotes the ith column of the histogram of $\mathbf{a}_n$.

Once Equation (11) is instantiated, we can respectively rewrite Equations (13) and (18) by:

$$s_{nm} = \frac{exp(-\frac{\|\mathcal{G}_{nm}(\mathbf{B},\mathbf{A})\|^2 + \Psi_{nm}(\mathbf{B},\mathbf{A})}{2\sigma^2})}{\sum_{i=1}^{N} exp(-\frac{\|\mathcal{G}_{im}(\mathbf{B},\mathbf{A})\|^2 + \Psi_{im}(\mathbf{B},\mathbf{A})}{2\sigma^2}) + \frac{2\pi\sigma^2\omega N}{M(1-\omega)}}, \tag{25}$$

and

$$Q = \frac{1}{2\sigma^2} \sum_{m=1}^{M} \sum_{n=1}^{N} s_{nm}(\|\mathcal{G}_{nm}(\overline{\mathbf{B}}, f(\mathbf{A}))\|^2 + \Psi_{nm}(\mathbf{B},\mathbf{A})) + N_{\mathbf{S}} log\sigma^2 + \frac{\mu}{2}\mathcal{R}(f); \tag{26}$$

thus, we can complete the M-step of EM algorithm for implementing the agricultural terrace image registration. The matrix form of Equation (26) to simplify the derivative is written as:

$$\begin{aligned} Q = &\frac{1}{2\sigma^2}(Tr(\mathcal{G}_{\overline{\mathbf{B}}}^T \mathbf{S}_{\overline{\mathbf{B}}} \mathcal{G}_{\overline{\mathbf{B}}}) - 2Tr(\mathcal{G}_{f(\mathbf{A})}^T \mathbf{S} \mathcal{G}_{\overline{\mathbf{B}}}) + Tr(\mathcal{G}_{f(\mathbf{A})}^T \mathbf{S}_{\mathbf{A}} \mathcal{G}_{f(\mathbf{A})}) - 2Tr(W^T \mathbf{G} \mathbf{S} \mathcal{G}_{\overline{\mathbf{B}}}) \\ &+ 2Tr(W^T \mathbf{G} \mathbf{S}_{\mathbf{A}} \mathcal{G}_{f(\mathbf{A})}) + Tr(W^T \mathbf{G} \mathbf{S}_{\mathbf{A}} \mathbf{G} W) + Tr(\Psi^T(\mathbf{B},\mathbf{A}) \mathbf{S} \Psi(\mathbf{B},\mathbf{A}))) + N_{\mathbf{S}} log\sigma^2 + \frac{\mu}{2} Tr(W^T \mathbf{G} W), \end{aligned} \tag{27}$$

where Tr denote trace operate, $\mathbf{S}_{\mathbf{A}} = dig(\mathbf{S1})$, $\mathbf{S}_{\mathbf{B}} = dig(\mathbf{S}^T\mathbf{1})$, $\mathbf{1}$ is a column vector with all ones. $\mathcal{G}_{\mathbf{P}} = \mathbf{P} + \mathbf{U}^T(\mathbf{P})\mathbf{P}$, where operator $\mathbf{U}(\mathbf{P})$ is defined by

$$\mathbf{U}_{ij}(\mathbf{P}) = \begin{cases} \eta_{ik} - K, & \overrightarrow{\mathbf{p}_i\mathbf{p}_j} \in \{\overrightarrow{\mathbf{p}_i\mathbf{p}_{i_k}}\}_{k=1}^{K}, \\ -K, & \overrightarrow{\mathbf{p}_i\mathbf{q}_j} \notin \{\overrightarrow{\mathbf{p}_i\mathbf{q}_{i_k}}\}_{k=1}^{K}, \end{cases}$$

with $\mathbf{P} = \{\mathbf{p}_i\}_{i=1}^{I}$ denoting a non-representational point set containing I points, i_k is the index of the kth neighbor of the ith point, K is the number of neighbors, and η_{ik} is the weight of LGS defined by Equation (9). The partial derivative of Equation (27) with respect to the parameter W is obtained by:

$$\frac{\partial Q}{\partial W} = -\frac{\mathbf{G}\mathbf{S}\mathcal{G}_{\overline{\mathbf{B}}}}{\sigma^2} + \frac{\mathbf{G}\mathbf{S}_{\mathbf{A}}\mathcal{G}_{\mathbf{A}}}{\sigma^2} + \frac{\mathbf{G}\mathbf{S}_{\mathbf{A}}\mathbf{G}W}{\sigma^2} + \mu\mathbf{G}W. \tag{28}$$

Setting Equation (28) to zero, the parameter W is obtained by:

$$W = (\mathbf{G}\mathbf{S}_{\mathbf{A}}\mathbf{G} + \mu\sigma^2\mathbf{G})^{-1}(\mathbf{G}\mathbf{S}\mathcal{G}_{\overline{\mathbf{B}}} - \mathbf{G}\mathbf{S}_{\mathbf{A}}\mathcal{G}_{\mathbf{A}}). \tag{29}$$

Parameter Setting

For evaluating image features, four groups of parameters—(i) R, the radius of the circular neighborhood in DRLBP; (ii) L, the number of the neighbors in DRLBP; (iii) τ, the window size controlled parameter in LT; and (iv) T_1 and T_2, two annealing parameters for LGS and LT—are used. We set $R = 1$, $L = 8$, $\tau = 10$, $T_1 = exp(-iter/10)$ and $T_2 = exp(-iter/50)$.

For registering extracted feature points, six parameters—(i) ω, outlier weighting parameter; (ii) α, a constant to control the spatial smoothness; (iii) σ^2, the equal isotropic covariances of MF; (iv) W, the parameter of point set transformation; (v) μ, the weighting parameter of regularization and (vi) $iter_{max}$, the max number of iteration—are used. We set $\omega = 0.7$, $\alpha = 2$ and $iter_{max} = 50$, and initialize W as a matrix with all zeros. Moreover, σ^2 and μ are first initialized by

$$\sigma^2 = \frac{NTr(\mathbf{A}^T\mathbf{A}) - 2(\mathbf{1A})(\mathbf{1B})^T + MTr(\mathbf{B}^T\mathbf{B})}{2NM}$$

and $\mu = 8$, then annealed as

$$\sigma^2 \leftarrow \frac{|Tr(\mathbf{A}^T\mathbf{S}_{\mathbf{A}}\mathbf{A}) - 2Tr(f(\mathbf{A})^T\mathbf{S}\overline{\mathbf{B}}) + Tr(\overline{\mathbf{B}}^T\mathbf{S}_{\mathbf{B}}\overline{\mathbf{B}})|}{2N_{\mathbf{S}}}$$

and

$$\mu \leftarrow \frac{(iter_{max}^4 - iter^4 + 1)^{\frac{1}{4}}}{iter_{max}} \times \mu,$$

respectively.

The pseudo-code of our feature points based agricultural terrace image registration is outlined in Algorithm 1.

Algorithm 1: Local texture feature and geometric structure feature guided agricultural terrace image registration

```
input     : Sensed agricultural terrace image $\mathbf{I}$ and reference agricultural terrace image $\mathbf{I}^t$
output    : Transformed point set $\hat{\mathbf{A}}$, correspondence matrix $\mathbf{S}$ and transformed image $\hat{\mathbf{I}}$
parameter : $R$, $L$, $\tau$, $T_1$, $T_2$, $\omega$, $\alpha$, $\sigma^2$, $W$, $\mu$ and $iter_{max}$

Image preprocessing:
    Graying image $\mathbf{I}$ and $\mathbf{I}^t$;
    Strengthen gray images by Equation (2);
    Expand the gray value distribution by Equation (3);
end
Select feature point set $\mathbf{A}$ and $\mathbf{B}$ from $\mathbf{I}_{new}$ and $\mathbf{I}^t_{new}$, respectively;
Initialize $\sigma^2$ and $W$;
Compute $\mathcal{LT}_{R,L}(\mathbf{A})$ and $\mathcal{LT}_{R,L}(\mathbf{B})$ in $\mathbf{I}$ and $\mathbf{I}^t$;
Construct Gaussian kernel $\mathbf{G}$;
repeat
    Correspondence estimation:
        Compute $\mathcal{LGS}(\mathbf{A})$ and $\mathcal{LGS}(\mathbf{B})$ by Equation (8), respectively;
        Compute the posterior probability matrix $\mathbf{S}$ by Equation (25);
        Compute the corresponding target point set $\overline{\mathbf{B}}$ by Equation (14);
    end
    Transformation estimation:
        Update the parameter $W$ by Equation (29);
        Update the source point set by $\mathbf{A} \leftarrow f(\mathbf{A})$;
    end
    Anneal $T_1$, $T_2$ and $\mu$;
    Anneal $\sigma^2$;
until reach max iteration number or Equation (26) is convergent;
Transformed point set $\hat{\mathbf{A}} = \mathbf{A}$ and correspondence matrix $\mathbf{S}$ are obtained in the final iteration;
Transformed agricultural terrace image $\hat{\mathbf{I}}$ is obtained by Equation (21).
```

2.6. Computational Complexity

The computational complexity of each part in our feature points based agricultural terrace image registration is as follows:

- Image preprocessing:
  time complexity: $\mathcal{O}(X_1Y_1 + X_2Y_2)$, space complexity: $\mathcal{O}(X_1Y_1 + X_2Y_2)$.
- Feature extraction:
  time complexity for feature points selection : $\mathcal{O}(X_1Y_1 + X_2Y_2)$, space complexity: $\mathcal{O}(N + M)$;
  time complexity for computing LT: $\mathcal{O}(NX_1^2Y_1^2 + MX_2^2Y_2^2)$, space complexity: $\mathcal{O}(NX_1Y_1 + MX_2Y_2)$;
  time complexity for computing LGS: $\mathcal{O}(N^2 + M^2)$, space complexity: $\mathcal{O}(N + M)$.
- Point set registration:
  time complexity for correspondence estimation: $\mathcal{O}(NM)$, space complexity: $\mathcal{O}(NM)$;
  time complexity for transformation estimation: $\mathcal{O}(N^3 + M^3)$, space complexity: $\mathcal{O}(N^2 + M^2 + NM)$.
- Image registration:
  time complexity: $\mathcal{O}(N^2 + X_1Y_1N + X_1Y_1 + M^2 + X_2Y_2M + X_2Y_2)$, space complexity: $\mathcal{O}(N^2 + X_1Y_1N + X_1Y_1 + M^2 + X_2Y_2M + X_2Y_2)$,

where X_1, X_2, Y_1, Y_2 are the widths and heights of the source image and sensed image, respectively; N, M are numbers of the feature points extracted from the source and sensed image. Overall, the time complexity of the proposed method is $\mathcal{O}(NX_1^2Y_1^2 + MX_2^2Y_2^2)$, and the space complexity is $\mathcal{O}(NX_1Y_1 + MX_2Y_2)$.

3. Experiments and Results

3.1. Experiments Design

CPD (coherent point drift) [29], GLMDTPS (global and local mixture distance with thin plate spline transformation) [30], SIFT (scale invariant feature transform) [38] and SURF (speeded-up robust features) [39], four state-of-the-art methods, are compared against our method in the following experiments. SIFT and SURF methods used the open source $VLFeat$ toolbox with the threshold 1 and the Matlab open source $OpenSURF$ function with the default setting, respectively. We design two series of experiments: (i) due to the employment of the same feature point sets, the quantitative comparison on feature point matching is carried out on CPD, GLMDTPS and our method using the precision ratio (PR) [40]; (ii) quantitative comparison and qualitative demonstration on image registration are carried out on all the methods using the root of mean square error (RMSE), mean absolute error (MAE) and standard deviation (SD). The experimental dataset includes 20 pairs multi-temporal (4–5 month time interval) and multi-viewpoint agricultural terrace images (500 × 300 pixels) captured from Longji and Yunhe terraces, China (see Figure 6). All agricultural terrace images were obtained by a small UAV (the DJI Phantom 4 Pro (SZ, China), store homepage: [41]) with a CMOS camera. The small UAV basically maintained the same flight height (around 50–70 m) for collecting multi-temporal images of the same locations, but appropriately changed the imaging perspective for generating different geometric distortions, and different overlapping degrees of image pairs. All experiments are tested on a PC with 2.60 GHz Intel CPU and 16 GB memory.

3.2. Evaluation Criterion

The PR is usually employed to estimate the accuracy of feature point matching and defined as

$$PR = \frac{TP}{TP + FP}, \tag{30}$$

where TP and FP denote the true positive and the false positive, respectively. The positive indicates inliers and the negative indicates outliers.

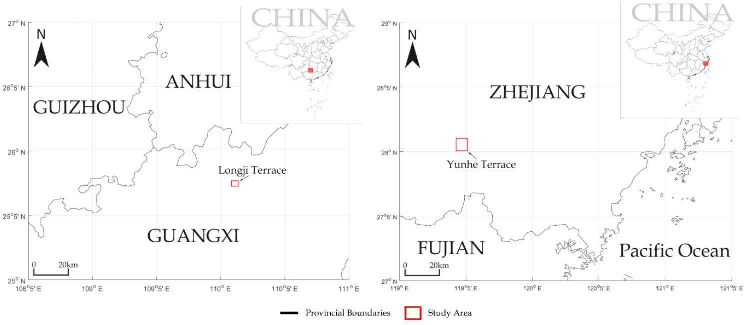


Figure 6. Location of Longji and Yunhe terraces. Longji terrace is located in Guangxi province, China (Longitude range: 110°08′65″E to 110°14′15″E; Latitude range: 25°72′51″N to 2°76′77″N.). Yunhe terrace is located in Zhejiang province, China (Longitude range: 11°42′73″E to 119°51′11″E; Latitude range: 28°00′58″N to 28°10′01″N).

The RMSE, MAE and SD are usually used to quantify image registration accuracy. We manually determine 20 pairs of landmarks between the sensed image and the reference image as ground truth, and all the landmarks are well-distributed and selected in the easily identified places around agricultural terraces. The related formulations and the definitions in statistics are as follows:

$$RMSE = \sqrt{\frac{1}{N^l} \sum_{n=1}^{N^l} \|a_n^l - b_n^l\|^2}, \tag{31}$$

$$MAE = \frac{\sum_{n=1}^{N^l} \|a_n^l - b_n^l\|}{N^l}, \tag{32}$$

$$SD = \sqrt{\frac{1}{N^l} \sum_{n=1}^{N^l} (d(a_n^l, b_n^l) - RMSE)^2}, \tag{33}$$

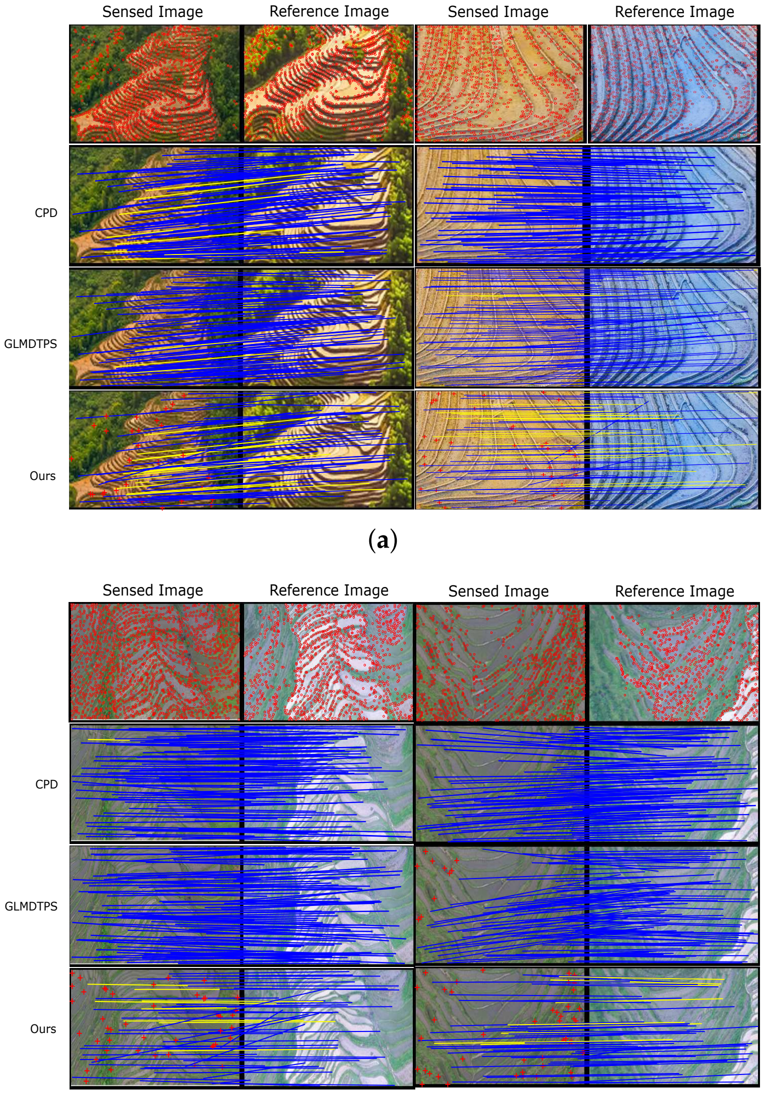
where a_n^l and b_n^l are the nth pair of corresponding landmarks in the sensed image and the reference image, respectively. N^l is the total number of selected landmarks, and the operator $d(\cdot,\ \cdot)$ denotes the distance.

3.3. Results of Feature Matching

All agricultural terrace image pairs contain viewpoint changes and were captured with a 4–5 month time interval. By using the proposed preprocessing method, each image pair has 624 to 1513 feature points. For the quantitative comparison and visual demonstration, we evenly selected 100 feature points with more than 50% outlier (negative) to data rate for calculating the PR and visualizing the matching results as shown in Table 1 and Figure 7. CPD and GLMDTPS gave poor performances since they estimated correspondences only using the Euclidean distance and the mixed geometric features, respectively, although CPD applied the motion coherent based geometric constraint to regularize the displacement field and GLMDTPS applied the annealing scheme to gradually change the transformation from rigid to non-rigid during registration. Our method gave the best matching performance in all image pairs since the local texture feature and the local geometric structure feature are combined and very complementary.

Table 1. Experimental results on series (i). Quantitative comparisons on the mean PR (precision ratio). Bold fonts indicate the best results. All units are in percentages. CPD (coherent point drift) denotes the method [29] and GLMDTPS (global and local mixture distance with thin plate spline transformation) denotes the method [30].

Method	CPD	GLMDTPS	Ours
PR	1.45%	10.43%	**63.1%**



(**a**)

(**b**)

Figure 7. Feature matching demonstrations on four typical agricultural terrace image pairs. In (**a**,**b**), the first to the fourth rows are: the image pairs, the results on feature matching of CPD (coherent point drift), GLMDTPS (global and local mixture distance with thin plate spline transformation), and ours, respectively. Red circles denote feature points extracted by our method. Blue lines indicate the false positive and the false negative, yellow lines indicate the true positive, and red crosses indicate the true negative.

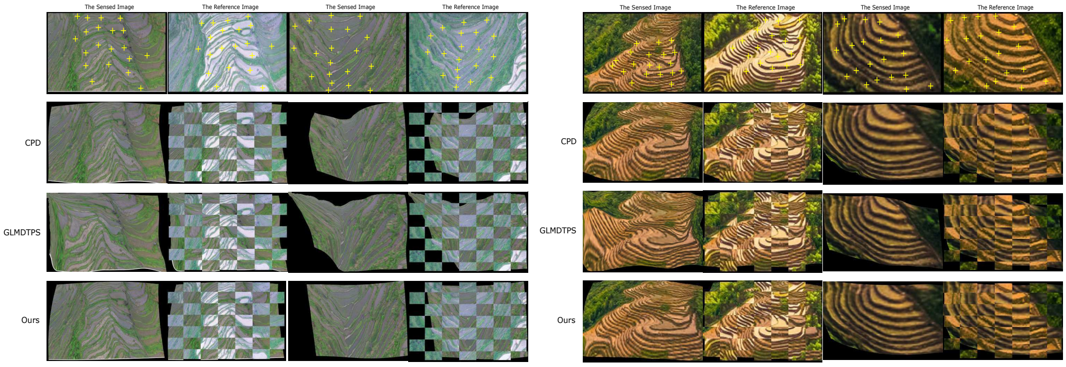
3.4. Results of Image Registration

For CPD, GLMDTPS and our method, all extracted feature points were used for image registration. Feature points for SIFT and SURF were extracted by their default setting. The quantitative comparison using the mean RMSE, MAE and SD are shown in Table 2. The transformed images and the checkboards in four typical image registration examples are shown in Figure 8. Registering images using sparse correspondence has the advantages of being potentially faster and easily maintains a high feature point matching ratio. However, the image registration accuracy might not be high since a desired image transformation with acceptable accuracy can only be established based upon more dense correspondence (e.g., more feature points). This can be explained since the image transformation is basically interpolated from correspondences of feature points, thereby an adequate number of correspondences play a crucial role in yielding detailed transformation.

In this experiment, SIFT and SURF, which extract a relative small number of feature points, failed all and 17 registrations, respectively, although SURF performed well in the other three registrations (see Table 2). The reason was that the image registrations with a small number of feature points are sensitive to mismatching. Although CPD and GLMDTPS employed the same number of feature points with our method, the geometric features used in CPD and GLMDTPS were sensitive to outliers and similar neighborhood structures in multi-temporal images. Therefore, they also gave the relatively poor registration accuracies (e.g., GLMDTPS failed five registrations). In our method, the local texture feature and the local geometric structure feature of terrace images are combined well to improve the feature description of points, while helping to reject outliers. The outliers were rejected in the point matching step, but used to yield detailed transformation in the image registration step. Therefore, our method gave the best registration performance.

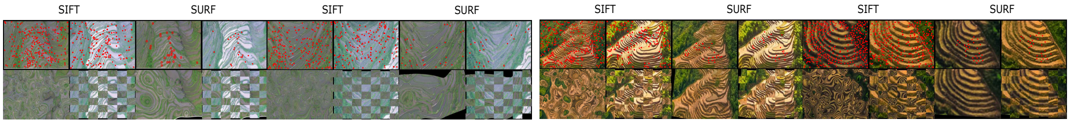
Table 2. Experimental results on series (ii). Quantitative comparisons on image registration measured using the mean RMSE (root of mean square error), MAE (mean absolute error) and SD (standard deviation) are carried out. $0 \leq \theta \leq 20$ denotes the number of the failed registrations (can not manually identify landmarks). Bold fonts indicate the best results, and all units are in pixels. CPD (coherent point drift) denotes the method [29], GLMDTPS (global and local mixture distance with thin plate spline transformation) denotes the method [30], SIFT (scale invariant feature transform) denotes the method [38] and SURF (speeded-up robust features) denotes the method [39].

Method	θ	**RMSE**	**MAE**	**SD**
CPD	0	41.78	25.18	50.54
GLMDTPS	5	21.35	6.95	24.51
SIFT	20	-	-	-
SURF	17	11.51	3.51	14.30
Ours	0	**29.95**	**10.70**	**37.89**



(**a**) (**b**)

Figure 8. *Cont.*



(**c**) (**d**)

Figure 8. Agricultural terrace image registration examples on four typical image pairs. In (**a**,**b**), the first to the fourth rows are: the image pairs, transformed images and checkboards built by CPD (coherent point drift), GLMDTPS (global and local mixture distance with thin plate spline transformation), and ours, respectively. Yellow crosses denote 20 pairs landmarks. In (**c**,**d**), the first rows show the image pairs and the second rows show the transformed images and the checkboards built by SIFT (scale invariant feature transform) and SURF (speeded-up robust features), respectively. Red circles denote the extracted feature points.

4. Conclusions

Due to soil erosion, illegal land occupations, agricultural land management practices and low purchasing power in China's southeast and southwest mountain area, smallholder farmers as well as local governments require a light-weight and low cost technology such as DJI Phantom UAVs to monitor their planting area in terraces. However, multi-temporal images of the same planting area captured by small UAVs have only visible image information and are always accompanied by viewpoint changes, image geometric distortions, low image overlapping, brightness changes and color changes such that the images may not be directly used for dynamic agricultural terrace monitoring. Thus, transforming multi-temporal images into one coordinate system is necessary in order to be able to subsequently compare or integrate planting area information for dynamic agricultural terrace monitoring.

In this work, we have presented a small UAV based multi-temporal image registration method. The proposed method first designed a guided image filtering to enhance terrace ridges in multi-temporal images, and a multi-feature descriptor was applied to combine the texture feature and the geometric structure feature of terrace images for improving the description of feature points and rejecting outliers. The multi-feature guided model then provided an accurate guiding for feature point set registration, and the feature points based image registration finally gave an accurate image registration. Experiments on 20 pairs of multi-temporal terrace images captured by a DJI Phantom 4 Pro demonstrated that our method gives the best registration performance, and outperforms four state-of-the-art methods. To fully realize the dynamic agricultural terrace monitoring, future work will focus on core decision rules of automatic change detection algorithms between the registered images. For registering other landscape elements, the image enhancement preprocessing, and multi-feature extraction and combination steps (as described in Sections 2.1 and 2.2), should focus on object-specific features that are invariant to various imaging perspectives, color and brightness changes.

Acknowledgments: The authors wish to thank David G. Lowe, Herbert Bay, Andriy Myronenko, Zhaoxia Liu and Yang Yang for providing their implementation source codes and test data sets. This greatly facilitated the comparison experiments. This work was supported by (i) the National Nature Science Foundation of China (41661080); (ii) the Scientific Research Foundation of Yunnan Provincial Department of Education (2017TYS045); (iii) the Doctoral Scientific Research Foundation of Yunnan Normal University (01000205020503065); and (iv) the National Undergraduate Training Program for Innovation and Entrepreneurship (201610681002).

Author Contributions: Yang Yang, Ziquan Wei and Kun Yang developed the method; Kun Yang and Yi Luo designed the data acquisition of agricultural terrace images by a small UAV; Mengya Li and Yifeng Han conceived and designed the experiments; Yang Yang, Ziquan Wei, Mengya Li and Yifeng Han performed the experiments and analyzed the data; Sim-Heng Ong and Ziquan Wei helped technology implementation of the method; Yang Yang and Ziquan Wei wrote the paper. All the authors reviewed and provided valuable comments for the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guo, Y.; Wu, Y.; Ju, Z.; Wang, J.; Zhao, L. Remote sensing image classification by the Chaos Genetic Algorithm in monitoring land use changes. *Math. Comput. Model.* **2010**, *51*, 1408–1416.
2. Małgorzata, W.; Piotr, M. Automatic relief classification versus expert and field based landform classification for the medium-altitude mountain range, the Sudetes, SW Poland. *Geomorphology* **2014**, *206*, 133–146.
3. Doxani, G.; Karantzalos, K.; Tsakiri-Strati, M. Monitoring urban changes based on scale-space filtering and object-oriented classification. *Int. J. Appl. Earth Obs. Geoinform.* **2012**, *15*, 38–48.
4. Müllerová, J.; Pergl, J.; Pyšek, P. Remote sensing as a tool for monitoring plant invasions: Testing the effects of data resolution and image classification approach on the detection of a model plant species *Heracleum mantegazzianum* (giant hogweed). *Int. J. Appl. Earth Obs. Geoinform.* **2013**, *25*, 55–65.
5. Drăguţ, L.; Blaschke, T. Automated classification of landform elements using object-based image analysis. *Geomorphology* **2006**, *81*, 330–344.
6. Ho, L.; Yamaguchi, Y.; Umitsu, M. Automated micro-landform classification by combination of satellite images and SRTM DEM. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 3058–3061.
7. Lu, D.; Mausel, P.; Brondízio, E.; Moran, E. Change detection techniques. *Int. J. Remote Sens.* **2004**, *25*, 2365–2401.
8. Prima, O.; Ayako, E.; Ryuzo, Y.; Takeyoshi, Y. Supervised landform classification of Northeast Honshu from DEM-derived thematic maps. *Geomorphology* **2006**, *78*, 373–386.
9. Camargo, F.; Almeida, C.; Florenzano, T.; Heipke, C.; Feitosa, R.; Costa, G. ASTER/Terra Imagery and a Multilevel Semantic Network for Semi-automated Classification of Landforms in a Subtropical Area. *Photogramm. Eng. Remote Sens.* **2011**, *11*, 619–629.
10. Karydas, C.G.; Sekuloska, T.; Sarakiotis, I. Fine scale mapping of agricultural landscape features to be used in environmental risk assessment in an olive cultivation area. *IASME Trans.* **2005**, *4*, 582–589.
11. Pradhan, B.; Chaudhari, A.; Adinarayana, J.; Buchroithner, M.F. Soil erosion assessment and its correlation with landslide events using remote sensing data and GIS: A case study at Penang Island, Malaysia. *Environ. Monit. Assess.* **2012**, *184*, 715–727.
12. Ventura, G.; Vilardo, G.; Terranova, C.; Sessa, E.B. Tracking and evolution of complex active landslides by multi-temporal airborne LiDAR data: The Montaguto landslide (Southern Italy). *Remote Sens. Environ.* **2011**, *115*, 3237–3248.
13. Martha, T.R.; Kerle, N.; van Westen, C.J.; Jetten, V.; Kumar, K.V. Object-oriented analysis of multi-temporal panchromatic images for creation of historical landslide inventories. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 105–119.
14. Bailly, J.S.; Levavasseur, F. Potential of linear features detection in a Mediterranean landscape from 3D VHR optical data: Application to terrace walls. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 7110–7113.
15. Wulder, M.A.; White, J.C.; Nelson, R.F.; Næsset, E.; Ørka, H.O.; Coops, N.C.; Hilker, T.; Bater, C.W.; Gobakken, T. Lidar sampling for large-area forest characterization: A review. *Remote Sens. Environ.* **2012**, *121*, 196–209.
16. Demoulin, A.; Bovy, B.; Rixhon, G.; Cornet, Y. An automated method to extract fluvial terraces from digital elevation models: The Vesdre valley, a case study in eastern Belgium. *Geomorphology* **2007**, *91*, 51–64.
17. Martínez-Casasnovas, J.A.; Ramos, M.C.; Cots-Folch, R. Influence of the EU CAP on terrain morphology and vineyard cultivation in the Priorat region of NE Spain. *Land Use Policy* **2010**, *27*, 11–21.
18. Del Val, M.; Iriarte, E.; Arriolabengoa, M.; Aranburu, A. An automated method to extract fluvial terraces from LiDAR based high resolution digital elevation models: The Oiartzun Valley, a case study in the Cantabrian margin. *Quat. Int.* **2015**, *364*, 35–43.
19. Li, Y.; Gong, J.; Wang, D.; An, L.; Li, R. Sloping farmland identification using hierarchical classification in the Xi-He region of China. *Int. J. Remote Sens.* **2013**, *34*, 545–562.
20. Gioia, D.; Bavusi, M.; Di Leo, P.; Giammatteo, T.; Schiattarella, M. A geoarchaeological study of the metaponto coastal belt, southern Italy, based on geomorphological mapping and gis-supported classification of landforms. *Geogr. Fis. Din. Quat.* **2016**, *39*, 137–148.

21. Yang, K.; Pan, A.; Yang, Y.; Zhang, S.; Ong, S.H.; Tang, H. Remote Sensing Image Registration Using Multiple Image Features. *Remote Sens.* **2017**, *9*, 581.
22. Diaz-Varela, R.; Zarco-Tejada, P.; Angileri, V.; Loudjani, P. Automatic identification of agricultural terraces through object-oriented analysis of very high resolution DSMs and multispectral imagery obtained from an unmanned aerial vehicle. *J. Environ. Manag.* **2014**, *134*, 117–126.
23. Deffontaines, B.; Chang, K.J.; Champenois, J.; Fruneau, B.; Pathier, E.; Hu, J.C.; Lu, S.T.; Liu, Y.C. Active interseismic shallow deformation of the Pingting terraces (Longitudinal Valley-Eastern Taiwan) from UAV high-resolution topographic data combined with InSAR time series. *Geomat. Nat. Hazards Risk* **2016**, *8*, 120–136.
24. He, K.; Sun, J.; Tang, X. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1397–1409.
25. Levin, A.; Lischinski, D.; Weiss, Y. A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 228–242.
26. Shi, J.; Tomasi, C. Good features to track. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'94), Seattle, WA, USA, 21–23 June 1994; pp. 593–600.
27. Serby, D.; Meier, E.; Van Gool, L. Probabilistic object tracking using multiple features. In Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), Cambridge, UK, 26 August 2004; Volume 2, pp. 184–187.
28. Mehta, R.; Egiazarian, K. Dominant rotated local binary patterns (DRLBP) for texture classification. *Pattern Recognit. Lett.* **2016**, *71*, 16–22.
29. Myronenko, A.; Song, X. Point set registration: Coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 2262–2275.
30. Yang, Y.; Ong, S.H.; Foong, K.W.C. A robust global and local mixture distance based non-rigid point set registration. *Pattern Recognit.* **2015**, *48*, 156–173.
31. Ma, J.; Qiu, W.; Zhao, J.; Ma, Y.; Yuille, A.L.; Tu, Z. Robust L_2E Estimation of Transformation for Non-Rigid Registration. *IEEE Trans. Signal Process.* **2015**, *63*, 1115–1129.
32. Ma, J.; Zhao, J.; Ma, Y.; Tian, J. Non-rigid visible and infrared face registration via regularized Gaussian fields criterion. *Pattern Recognit.* **2015**, *48*, 772–784.
33. Ma, J.; Zhao, J.; Tian, J.; Yuille, A.L.; Tu, Z. Robust point matching via vector field consensus. *IEEE Trans. Image Process.* **2014**, *23*, 1706–1721.
34. Tikhonov, A.N.; Arsenin, V.Y. *Solutions of Ill-Posed Problems*; VH Winston & Sons: Washington, DC, USA, 1977.
35. Chen, Z.; Haykin, S. On different facets of regularization theory. *Neural Comput.* **2002**, *14*, 2791–2846.
36. Schölkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.
37. Ma, J.; Zhao, J.; Tian, J.; Bai, X.; Tu, Z. Regularized vector field learning with sparse approximation for mismatch removal. *Pattern Recognit.* **2013**, *46*, 3519–3532.
38. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2, pp. 1150–1157.
39. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359.
40. Jian, B.; Vemuri, B.C. Robust Point Set Registration Using Gaussian Mixture Models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1633–1645.
41. The Store Homepage of the DJI Phantom 4 Pro. Available online: http://store.dji.com/product/phantom-4-pro/ (accessed on 30 August 2017).