remote sensing

MDPI

Article

Generic and Automatic Markov Random Field-Based Registration for Multimodal Remote Sensing Image Using Grayscale and Gradient Information

Li Yan, Ziqi Wang *, Yi Liu * and Zhiyun Ye

School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China;
lyan@sgg.whu.edu.cn (L.Y.); yezhiyun1989@126.com (Z.Y.)
* Correspondence: zqwang0531@whu.edu.cn (Z.W.); yliu@sgg.whu.edu.cn (Y.L.)

Received: 2 July 2018; Accepted: 2 August 2018; Published: 5 August 2018

Abstract: The automatic image registration serves as a technical prerequisite for multimodal remote sensing image fusion. Meanwhile, it is also the technical basis for change detection, image stitching and target recognition. The demands of subpixel level registration accuracy can be rarely satisfied with a multimodal image registration method based on feature matching. In light of this, we propose a Generic and automatic Markov Random Field (MRF)-based registration framework of multimodal image using grayscale and gradient information. The proposed approach performs non-rigid registration and formulates an MRF model while grayscale and gradient statistical information of a multimodal image is employed for the evaluation of similarity while the spatial weighting function is optimized simultaneously. Besides, the value space is discretized to improve the convergence speed. The developed automatic approach was validated both qualitatively and quantitatively, demonstrating its potential for a variety of multimodal remote sensing datasets and scenes. As for the registration accuracy, the average target registration error of the proposed framework is less than 1 pixel, while the maximum displacement error is less than 1 pixel. Compared with the polynomial model registration based on manual selection, the registration accuracy has been significantly improved. In the meantime, the proposed approach had the partial applicability for the multimodal image registration of large deformation scenes. It is also proved that the proposed registration framework using grayscale and gradient information outperforms the MRF-based registration using only grayscale information and only gradient information while the proposed registration framework using Gaussian function as spatial weighting function is superior to that using distance inverse weight method.

Keywords: multimodal image; Markov Random Field; Grayscale and Gradient Information; spatial weighting function; non-rigid registration

1. Introduction

Multimodal remote sensing images can be applied to compensate the deficiencies of the single image source by increasing the amount of the image information. The generic and automatic registration of multimodal remote sensing image is the necessary step of point cloud coloring, image fusion, image stitching and mosaic, target recognition and change detection (e.g., change detection based on two heterogeneous images acquired by optical sensors and radars on different dates [1]). And it is of fundamental importance for numerous emerging geospatial environmental and engineering applications (e.g., geometric correction of SAR image using the optical image, band-to-band image registration developed for the High-Precision Telescope of microsatellite remote sensing [2], the spatial registration of point/line-scan hyperspectral sensor measurements in-water hyperspectral imaging [3]).

Therefore, multimodal image registration technology is the major obstacle impeding the improvement of the accuracy and effectiveness of various problems [4]. Due to the various imaging mechanisms of different sensors and the disparate time, angle and environment of image acquisition, there are still many challenges in the field of multimodal images registration. The realization of high-precision and automatic registration technology is even more difficult, especially in some cases where there are significant differences between remote sensing data category (e.g., optical and SAR images, point cloud depth map and panoramic images) and band (e.g., visible light and medium wave infrared image, near infrared and multi-spectral images).

Traditional remote sensing image registration methods adopt the manual selection of control points to solve geometric transformation relationships to achieve pixel-by-pixel alignment between images, which are mainly constrained in terms of two aspects, including the heavy workload and low efficiency and automation. At present, the automatic registration technology for multimodal remote sensing images has been extensively studied and is one of the research hotspots in the field of image processing.

For multimodal remote sensing images, the grayscale characteristic is no longer a linear relationship. Neither is it even a non-function change generally with statistical correlations and geometric similarities in the gray relations between images. Meanwhile, the non-uniform deformation would occur during acquisition of multimodal images. As what is shown in Figure 1, errors are in pixels and have been calculated based on manually denoted homonymy points. It sometimes would be ignored that numerous deformation properties of multimodal images are non-rigid and non-linear.

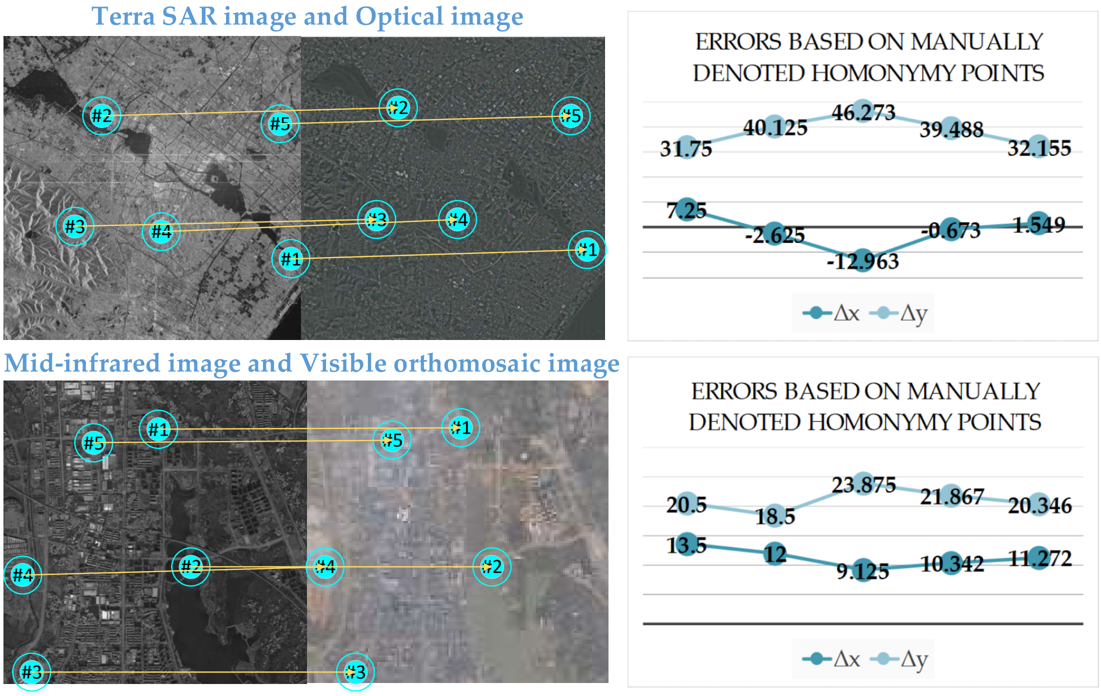


Figure 1. Non-uniform deformation of multimodal remote sensing image.

Image registration has been comprehensively studied through many approaches being proposed. Due to the weak grayscale correlation of multimodal images, simple, gray-based registration methods (e.g., the correlation function [5]) would lead to inaccurate registration. Feature descriptors, although it has great advantages in homologous image registration, do not perform with the same robustness in multimodal remote sensing datasets [6] while these methods (e.g., contour feature [7,8], SUSAN [9], SIFT [10] phase congruency based feature [11,12], extended SURF [13]) are faced with the challenges of repetitive feature detection and modal-invariant feature description. Most of them are still exploratory or are only suitable for specific occasions or certain types of multimodal images.

Due to the statistical correlation of grayscale relationships and the similarity of geometric structures between multimodal remote sensing images, the regional registration method based on statistical dependence or image structure consistency is suitable for multimodal image registration. The following is a summary of the current research status of this type of registration method.

The regional registration methods based on statistical dependency outperforms the previous methods. The well-known mutual information [14–16] mainly uses the statistical correlation of gray features between images, which is employed for medical image registration first and the multimodal image registration afterwards. The normalized mutual information [17] can be applied to overcome the shortcoming of mutual information sensitivity to image overlap. High-dimensional mutual information [18], regional mutual information [19], feature mutual information [20] have high dimension, complex calculation and other defects. Cross-Cumulative Residual Entropy (CCRE) [21] introduces cumulative residual entropy into mutual information with strong anti-noise ability and stability. Besides, wavelet transform [22] is applied to registration methods based on mutual information to realize the registration of infrared and optical images from coarse to fine. It is difficult for these methods to achieve ideal results in terms of time consumption or registration accuracy.

Generally, the basic structures of ground objects can be preserved in the multimodal image. These structures share certain similarities and invariances. Image structure information can generally be represented by gradients, edges and self-similarity. Yan et al. [23] proposed the gradient consistency operator based on the norm-weighted angle between gradient vectors and applied it to the registration between medium-infrared and visible images. Heinrich et al. [24] proposed Modality Independent Neighborhood Descriptor (MIND), which enjoys obvious advantages compared with the anti-noise characteristics and stability. However, it involves massive calculation. Ye et al. [25] construct a shape similarity measurement using local self-similarity and normalized correlation coefficients to for multispectral remote sensing image registration. These methods are only suitable for certain multimodal images or have low computational efficiency.

Another type of registration methods is to combines image structure information and mutual information. Among them, the most popular research is the combination of gradient feature information and mutual information for heterogeneous image registration [26–29], which is widely applied to medical image registration research. But in the field of remote sensing image registration and application, these methods are still immature and further research is needed.

In order to register multimodal remote sensing images, various registration methods have been proposed through these decades. A generic and automated registration framework based on Markov Random Field (MRF) [30,31] has been successfully utilized for multimodal remote sensing image registration. The discrete optimization setting along with the introduced data-specific energy terms form a modular approach with respect to the similarity criterion, which endows the fully exploitation of the spectral properties of multimodal remote sensing datasets.

However, the registration accuracy of the registration framework based on Markov Random Field (MRF) cannot reach the subpixel level. Inspired by the aforementioned work, we propose a common, automatic registration framework of the multimodal image to satisfy the standards of the desired registration accuracy. The proposed approach performs non-rigid registration, formulates a Markov Random Field (MRF) model while grayscale and gradient statistical information of multimodal image is employed for the evaluation of the similarity and the spatial weighting function is optimized simultaneously. Deformable registration satisfies differential homeomorphism, which can better maintain the topology of the image [32]. The value space is discretized to improve the convergence speed. In terms of spatial accuracy, the average registration error is less than 1 pixel while the maximum registration error is less than 1 pixel. The major contributions of this paper are listed below:

(1) We propose a generic and automatic registration framework for multiple multimodal remote sensing images which significantly improves registration accuracy.

(2) Compared with other methods, the proposed approach considers both the grayscale and gradient statistical information of images, which is used for similarity measures of data-specific energy term, hence to possess stronger robustness for the registration of multiple heterogeneous remote sensing images and target scenes.

(3) It can be noted that the contribution of the spatial weight function to the registration energy should be considered. We employed the Gaussian function method, which can overcome the discontinuity of the spatial weight function and hence to further improve registration accuracy.

The rest of the paper is organized as follows: In Section 2, we describe the overall process and details of the proposed generic and automatic MRF-based registration for multimodal remote sensing image using grayscale and gradient Information. Section 3 presents the experiments and results of four multimodal datasets to demonstrate the effectiveness both qualitatively and quantitatively. In Section 4, a discussion about the proposed method with the accuracy improvement and the influence of parameter analysis is conducted. Section 5 presents the conclusions and suggestions for the future work.

2. Materials and Methods

The multimodal image registration framework considering the gray and gradient information is the association of Markov Random Field (MRF) and B-spline interpolation. The basic framework of the framework is the Markov Random Field (MRF) model, which can be used to estimate the optimal transformation by minimizing the energy function. Firstly, the image is meshed to form a uniform control grid. Considering the gray and gradient information of the image for cost calculation, the optimal solution (Fast PD algorithm [33–35], used drop library) is used to obtain the label of MRF (the displacement of the control grid node). Then the B-spline interpolation strategy is used for the displacement calculation of each pixel point to control image deformation and perform registration.

In the specific case of image registration, the MRF model is defined as follows [30]: the node (the spatial position x_p) corresponds to the control point in the uniform B-spline grid; For each node, there is a set of discrete labels L; The discrete labels set L corresponds to the quantization of the solution space, representing the allowable discrete displacement. The number of labels set is $4n + 1$ while the uniform sampling number along the x, y and *diagonal* lines is n. The random variables would correspond with the displacement of the control points. The optimized energy function consists of two terms: the data term measures the data (the source image and the gradient image) likelihood of applying all allowed displacements to each random variable through the use of unary potentials D and the regular term penalizes non-desirable interactions between the random variables through the use of pairwise potentials P and introduces the prior knowledge of the smoothing constraints of the deformation field. Besides, λ is a scalar value used to evaluate the influence of the regular term. The goal of image registration is to assign an optimal label to every control grid node, so that the following energy is minimized:

$$E_{MRF}(l) = \underbrace{\sum_{p \in V} D_p(l_p)}_{data\ term} + \underbrace{\lambda \sum_{(pq) \in \varepsilon} P_{pq}(l_p, l_q)}_{regularisation\ term}; \quad (1)$$

At the same time, the framework was implemented in this paper coupled with the pyramidal representation of the images and a multi-scale approach for the deformation model. The Gaussian pyramids could reduce the computational cost while the multi-scale approach for the deformation model can increase the resolution by halving the interval of the control points. Therefore, we can gradually improve the registration results and restore the larger displacement. For every grid resolution level under every pyramid level, an iterative scheme was used in order to enhance the efficiency of discrete labels. It is necessary to keep a reasonable label set space and optimize the labels at each iteration. Meanwhile, the displacements of grid nodes are used to improve the registration image to capture smaller displacements. The registration process is summarized in Figure 2.

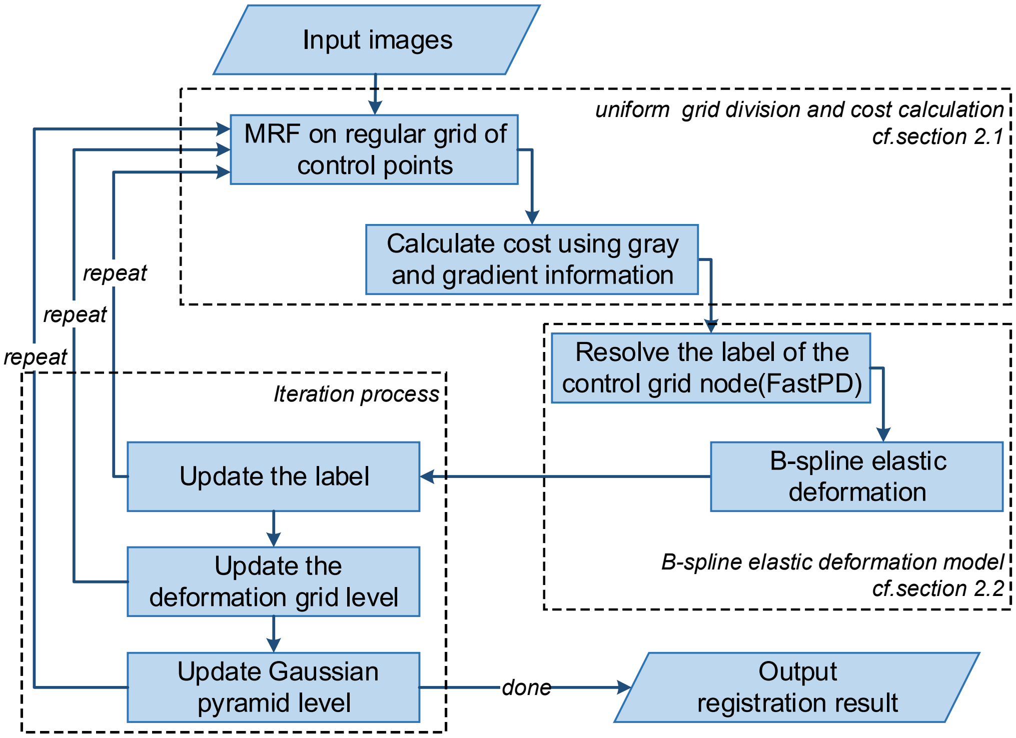

Input images
uniform grid division and cost calculation
cf.section 2.1
MRF on regular grid of control points
Calculate cost using gray and gradient information
repeat
repeat
repeat
Resolve the label of the control grid node(FastPD)
B-spline elastic deformation
B-spline elastic deformation model
cf.section 2.2
Iteration process
Update the label
Update the deformation grid level
Update Gaussian pyramid level
done
Output registration result

Figure 2. Overview of presented approach for multimodal image registration.

2.1. Uniform Grid Division of Control Point and Cost Calculation

The registration model we proposed in this paper is a nonparametric registration model based on MRF. The image was meshed to form a uniform control point grid G: $M \times N$ (M and N are significantly smaller than the image size; the interval between the grid nodes is δ). The cost calculation of regular grid codes based on the MRF model is composed by two parts, namely the grayscale data term and the gradient date term, which are defined by Equation (2). We employed a block-matching similarity criterion [30] to evaluate the maximum likelihood of applying all permissible displacements to each random variable, where the similarity measure is performed with the block image and its gradient image centered on each control grid node:

$$\sum_{p \in V} D_p(l_p) = \sum_{p \in V} U_p(l_p) + \omega \sum_{p \in V} G_p(l_p) \tag{2}$$

$$U_p(l_p) = \int_{\Omega} \hat{\eta}(\|x - p\|)\rho\left(I_S \circ d^{l_p}, I_T\right)dx \tag{3}$$

$$G_p(l_p) = \int_{\Omega} \hat{\eta}(\|x - p\|)\rho\left(G_S \circ d^{l_p}, G_T\right)dx \tag{4}$$

where I_S, G_S denote the original image and its corresponding gradient image. Accordingly, I_T, G_T refer to the target image and its gradient image. ω represents the gradient statistic weight, which decreases corresponding with the increase of the number of the iterations. η represents the block weighting function around the control point p, ranging from 0 to 1. The closer the space to the central control point is, the greater the impact is $\| \cdot \|$ denotes Euclidean norm. Normal distribution is obviously a desirable weight distribution model. The center point p is taken as the origin while the rest of the points are measured based on their positions on the normal curve. The density function of normal distribution is represented by Gaussian function:

$$\hat{\eta} = \exp(-(\|x - p\|/b)^2) \tag{5}$$

where b is the width parameter of the Gaussian function, which can be used to control the radial range.

In this paper, the gradient information of the multimodal image is extracted by the morphological gradient. The basic operation (e.g., morphological corrosion, expansion, open operation and closed operation) is used for image processing. The morphological gradient operator is formed correspondingly. Common morphological gradient operators can be calculated as follows:

$$G_1(f(x,y)) = (f \oplus B)(x,y) - f(x,y); \tag{6}$$

$$G_2(f(x,y)) = f(x,y) - (f\Theta B)(x,y); \tag{7}$$

$$G_3(f(x,y)) = (f \oplus B)(x,y) - (f\Theta B)(x,y) \tag{8}$$

The morphological gradient is the difference between the expansion graph and the corrosion graph. It can detect the edge of the image and extract the feature information and hence can obviously reduce the calculation load of the similarity measure and also improve the local extreme-value problem.

The similarity measure ρ can use mutual information that is insensitive to changes in light. The normalized mutual information of two images can be used for similarity measure calculation based on block matching strategy, which can reflect the degree of mutual information between them through their entropy and joint entropy:

$$nmi_{I_1,I_2}(i,k) = \frac{h_{I_1}(i) + h_{I_2}(k)}{h_{I_1,I_2}(i,k)} \tag{9}$$

where h_{I_1,I_2} can be calculated from the joint probability distribution of the corresponding gray scale [36]. The number of corresponding pixels is n. Parzen estimates [37] is used with the convolution of 2D Gaussian (represented by $\otimes g(i,k)$):

$$h_{I_1,I_2}(i,k) = -\frac{1}{n}\log(P_{I_1,I_2}(i,k) \otimes g(i,k)) \otimes g(i,k) \tag{10}$$

Accordingly, the calculation of h_{I_1}, h_{I_2} can be similar to h_{I_1,I_2}:

$$h_I(i) = -\frac{1}{n}\log(P_I(i) \otimes g(i)) \otimes g(i) \tag{11}$$

For the regularization term, the priori constraint of the smoothing deformation field is introduced. This indicates that the displacement field of the control grid nodes is assumed to be smooth while the spatially close variables p and q should be assigned to have similar labels. We employ a simple strategy that is based on the vector differences between candidate labels normalized by grid distance δ:

$$P_{pq}(l_p, l_q) = \frac{\|l_p - l_q\|}{\delta} \tag{12}$$

It is worth to note that, the cost calculation needs to use the grayscale and gradient information of the multimodal image based on normalization mutual information (NMI). At the same time, the block matching strategy is adopted for multimodal remote sensing datasets, which allows the local difference evaluation between the images to be registered while the computational efficiency can be improved. Cost calculation visualization flow chart shown in Figure 3.

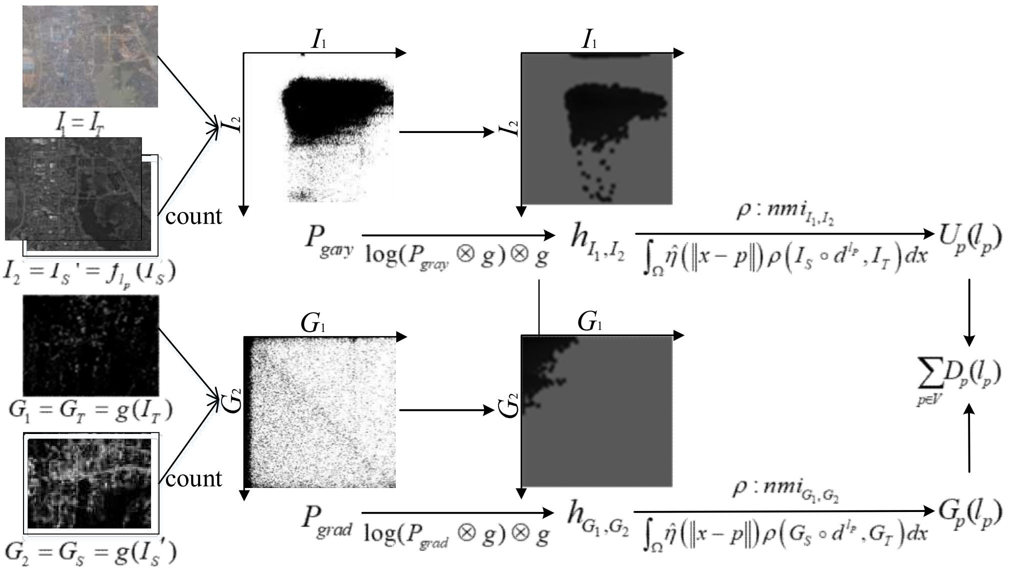

I_1
I_2
count
$I_1 = I_T$
$I_2 = I_S' = f_{l_p}(I_S)$
P_{gray}
$\log(P_{gray} \otimes g) \otimes g$
h_{I_1,I_2}
$\rho : nmi_{I_1,I_2}$
$\int_\Omega \hat{\eta}(\|x-p\|)\rho(I_S \circ d^{l_p}, I_T)dx$
$U_p(l_p)$
$\sum_{p \in V} D_p(l_p)$
G_1
G_2
$G_1 = G_T = g(I_T)$
$G_2 = G_S = g(I_S')$
P_{grad}
$\log(P_{grad} \otimes g) \otimes g$
h_{G_1,G_2}
$\rho : nmi_{G_1,G_2}$
$\int_\Omega \hat{\eta}(\|x-p\|)\rho(G_S \circ d^{l_p}, G_T)dx$
$G_p(l_p)$

Figure 3. Overview of cost calculation visualization. Values are scaled linearly for visualization. Darker points have larger values than brighter points.

2.2. B-Spline Elastic Deformation Model

By optimizing the MRF model, we can solve the labels of controls grid nodes. The elastic deformation model of B-spline based on the grid can provide one-to-one and reversible conversion. The basic idea of the deformation model is to deform the underlying image through the control of the nodes of the grid and calculate their influence in the rest of the image domain via an interpretation strategy:

$$T(x) = x + \sum_{i=1}^{K} \sum_{j=1}^{L} \eta_{ij}(x) d_{ij} \tag{13}$$

where $T(x)$ is the target image; d is the displacement of the control point ij; η corresponds to the interpolation or weighting function that determines the effect of the control point ij on the image point x. The closer the image point is, the greater the effect of the control point would be.

A uniform cubic B-spline function [38] is used by the interpolation strategy. The $(m + 3) \times (n + 3)$ control grid is used to define the $2D$ uniform cubic B-spline function. The function F_2 consists of $m \times n$ $2D$ patches, each of which is determined by 4 × 4 control points. Without loss of generality, a 2D uniform cubic B-spline function is represented by a patch f_2, which is defined by:

$$f_2(u, v) = (x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} B_i(u) B_j(v) \phi_{ij} \tag{14}$$

where, $0 \leq u, v \leq 1$, B_0, B_1, B_2 and B_3 is the basic function for the uniform cubic B-spline, ϕ_{ij} denotes to the displacement of the control point.

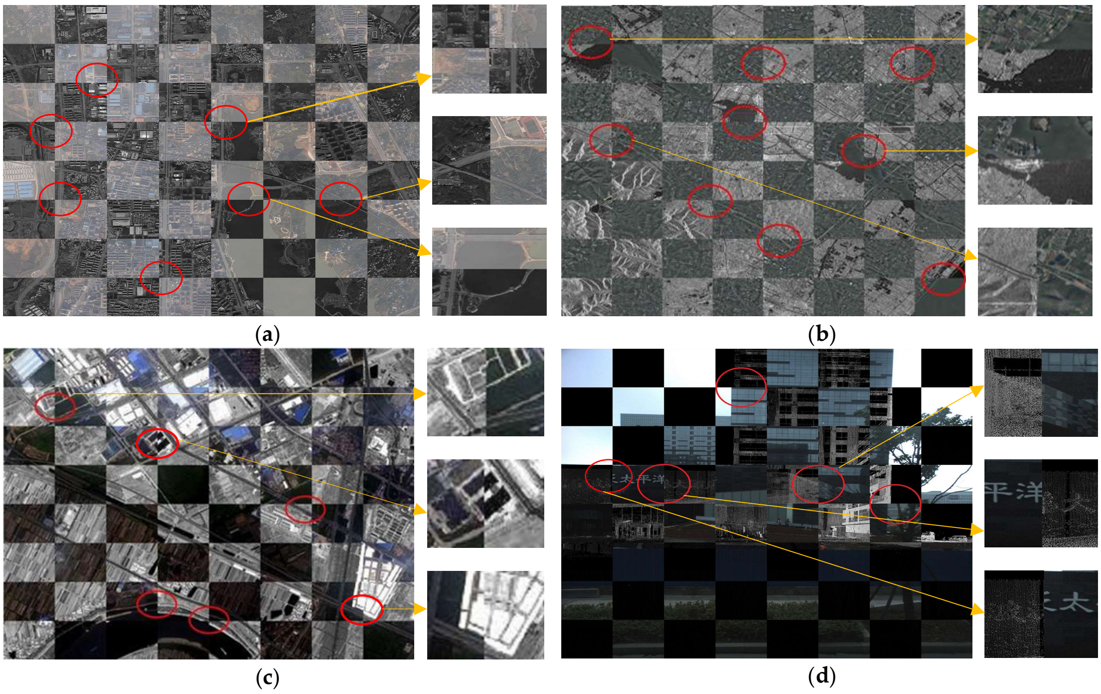
3. Experiments and Results

3.1. Descriptions of Experimental Data

Four remote sensing datasets of different scenes and applications have been used to evaluate our method. Several specific data of the four datasets are available in Table 1, shown in Figure 4.

Table 1. The multimodal datasets.

No.	Image Type	Image Size	Image Introduction	Maximum Parallax/Average Parallax
1	Mid-infrared image Visible orthomosaic image	2048 * 2048 2048 * 2048	Aerial image in Daye city of Hubei province	13 pixels/17 pixels
2	Terra SAR image Optical image	2048 * 2048 2048 * 2048	Changzhou City, Jiangsu Province, satellite survey area	48 pixels/23 pixels
3	Worldview-3 near infrared image Worldview-3 multispectral image	1536 * 2048 1536 * 2048	Wuhan City, Hubei Province, East and West Lake survey satellite imagery	29 pixels/10 pixels
4	Monolithic cloud point depth image Monolithic visible image	2048 * 2048 2048 * 2048	Wuhan City, vehicle-borne mobile image	21 pixels/17 pixels



(**a**) (**b**) (**c**) (**d**)

Figure 4. Checkerboard visualization results and local graphs of multimodal images before registration. The images were superimposed together, in the form of alternate checkerboard before registration. Red circles demonstrate the non-aligned area of the checkerboard block edge. For each group of multimodal images, three non-aligned areas were chosen for local amplification. The above row illustrates the raw unregistered image data: (**a**) Mid-infrared image and visible orthomosaic image; (**b**) Terra SAR image and optical image; (**c**) Worldview-3 near infrared image and multispectral image; (**d**) Monolithic cloud point depth image and visible image.

3.2. Multimodal Image Registration

3.2.1. Experimental Setup

There were four multimodal datasets being used in our experiment. There are a number of factors that should be considered when we design the following experiment setup, including three Gaussian pyramids, three deformation grids and the initial control point grid spacing δ, which is 64 pixels. The other two levels are set to 32 and 16 pixels, respectively. For each level of the pyramid network at all levels, a 5-iterative scheme is used, which applies the normalized mutual information as a similarity measure and the Gaussian function is used as the spatial weight function. Meanwhile, the morphological gradient is used to retain the edge of the ground object contour. The label set is 41 (along with the x, y axis and *diagonal* uniform sampling of 10 labels plus the origin to build the label set). In the first iteration, the maximum sample displacement equals to $0.4 \times \delta$, which satisfies the differential homeomorphism requirements and ensures that the topology of the image would not change [32]. According to the following iterations, the maximum sample shift corresponds to 0.67 of the maximum displacement of the previous iteration. The initial value of ω is 1. Corresponding with the increase of the number of iterations, the weight of the gradient statistics tends to zero; λ is the experience of choice. The patch size equals to $2\delta \times 2\delta$.

The arguments that have to be provided for running the FastPD algorithm are as follows: the *numpoints* is the number of grid nodes; the *numlabels* is 41; the *numpairs* is the number of MRF edges; the *max_iters* is 100; the *lcosts* and *wcosts* are obtained by cost calculation.

The optimal parameter settings for four multimodal datasets are shown in Table 2.

Table 2. The optimal parameter settings for four multimodal datasets.

Multimodal Dataset	Optimal Parameter Settings
Mid-infrared image and visible orthomosaic image	λ = 800, η = Gaussian function method.
Terra SAR image and optical image	λ = 400, η = Gaussian function method.
Worldview-3 near infrared image and multispectral image	λ = 1600, η = Gaussian function method.
Monolithic cloud point depth image and visible image	λ = 800, η = Gaussian function method.

In order to evaluate the effectiveness of the proposed registration framework, the experimental results are compared with three methods, including (1) polynomial registration model based on manually selection; (2) SIFT-based automatic registration [39]; (3) MRF-based registration framework using grayscale and gradient information.

The computer environment is based on a personal computer in Wuhan, Hubei Province, China with an Inter(R) Core(TM) i5-2320 processor, 8G memory without using the GPU for calculating the computational time. As for the computational efficiency of the developed registration framework, the calculation of four multimodal datasets takes about 7–8 minutes.

3.2.2. Experimental Results and Evaluation

The evaluation of the developed registration algorithm was performed both qualitatively and quantitatively. Various checkerboard visualization figures were closely reviewed along with the overlap images for the qualitatively evaluation, demonstrating that the unregistered and registered image and the regions of interest on the checkerboards were selected to local zoom for comparative analysis.

As it be seen in Figure 5, the polynomial registration model based on manually selection led to inaccurate results. Most of checkerboard lattice edge of multimodal image overlay after the registration had not been aligned.

As it can be seen from Figure 6, due to the differences of the grayscale characteristics and also the geometrical structure of heterogeneous remote sensing images, the SIFT-based automatic registration led to error results and was only suitable for Worldview-3 near infrared image and multispectral image registration.

As it can be seen from Figure 7, each block edge of checkerboard visualization has been aligned already, demonstrating the effectiveness of the proposed algorithm for multimodal data.

For quantitative evaluation, errors are in pixel level and have been calculated based on homonymy points, which are manually denoted in all registered images and reference images. The calculated registration errors (the dx, dy displacements along the axis and the distance D, in pixels) are given in Table 3.

Regarding the registration accuracy, the proposed algorithm using grayscale and gradient information outperforms other two methods. The average target registration errors of four remote sensing datasets are less than 1 pixel, while the maximum displacement errors are less than 1 pixel.

Above all, the proposed approach is optimal which has the best visualization results and meets the subpixel-level registration accuracy needs.

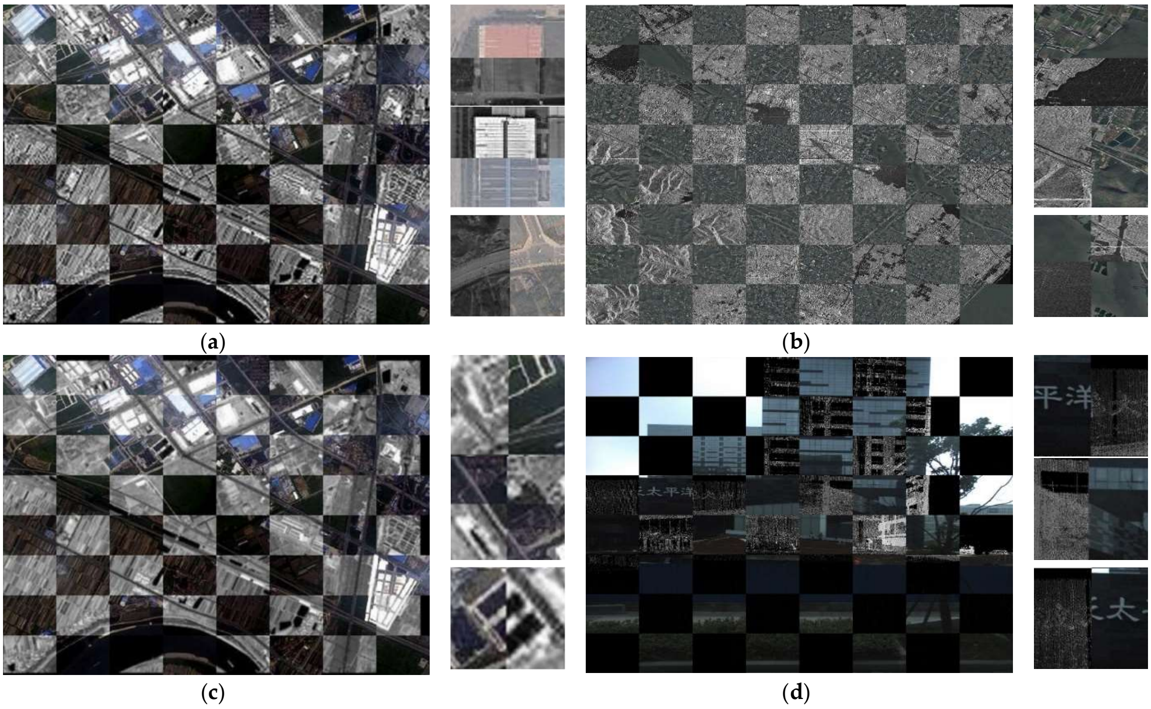

(**a**) (**b**)
(**c**) (**d**)

Figure 5. Checkerboard visualization for the qualitative evaluation of the polynomial registration model based on manually selection: (**a**) Mid-infrared image and visible orthomosaic image; (**b**) Terra SAR image and optical image; (**c**) Worldview-3 near infrared image and multispectral image; (**d**) Monolithic cloud point depth image and visible image.

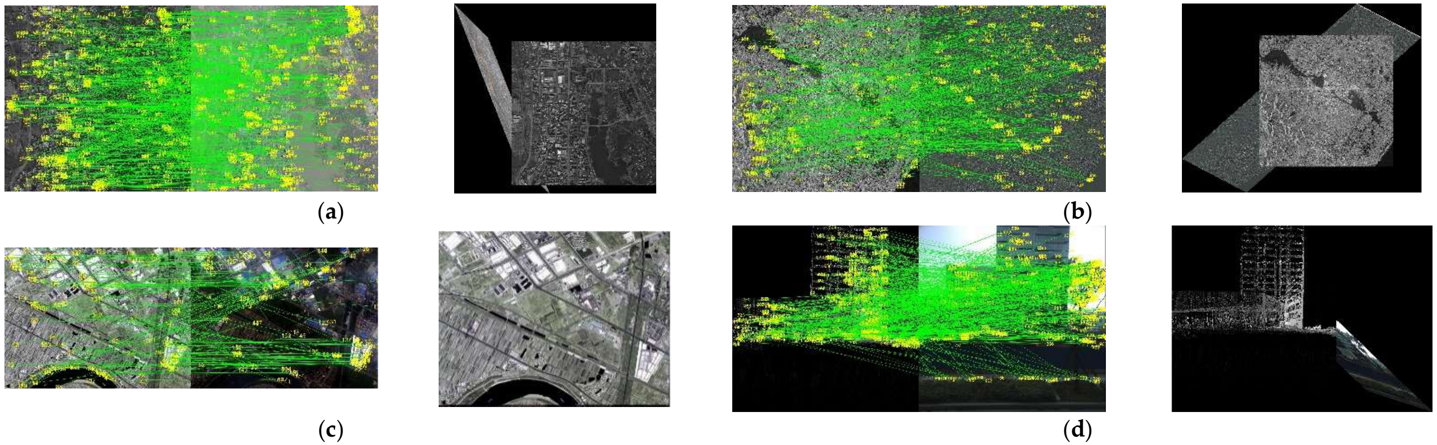


Figure 6. Keypoint matching and SIFT-based automatic registration results of four multimodal remote sensing datasets: (**a**) Mid-infrared image and visible orthomosaic image; (**b**) Terra SAR image and optical image; (**c**) Worldview-3 near infrared image and multispectral image; (**d**) Monolithic cloud point depth image and visible image.

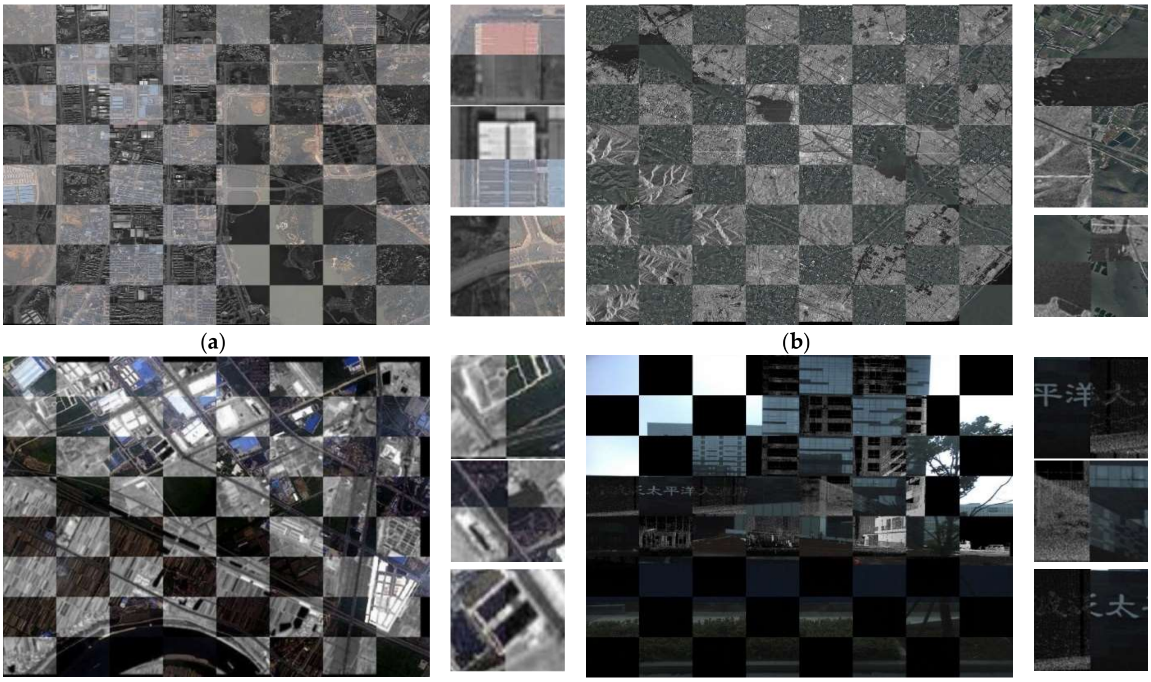


Figure 7. Checkerboard visualization and local zoom graph of the interest regions on the checkerboard visualization for the qualitative evaluation of the proposed multimodal registration framework: (**a**) Mid-infrared image and visible orthomosaic image; (**b**) Terra SAR image and optical image; (**c**) Worldview-3 near infrared image and multispectral image; (**d**) Monolithic cloud point depth image and visible image.

Table 3. Quantitative evaluation results after the application of the proposed multimodal registration framework.

Multimodal Dataset	Experimental Method	*dx* (pixels)	*dy* (pixels)	*D* (pixels)	*dx_max* (pixels)	*dy_max* (pixels)
Mid-infrared image and visible orthomosaic image	Manual selection + Polynomial Model	2.658	3.345	4.272	3.875	4.961
	Grayscale and gradient information + MRF	0.298	0.548	0.624	0.750	0.950
Terra SAR image and optical image	Manual selection + Polynomial Model	5.771	−8.961	10.658	6.520	−9.414
	Grayscale and gradient information + MRF	0.521	0.565	0.769	0.754	0.890
Worldview-3 near infrared image and multispectral image	Manual selection + Polynomial Model	3.107	4.251	5.265	4.268	6.632
	Grayscale and gradient information + MRF	0.344	0.406	0.532	0.537	0.562
Monolithic cloud point depth image and visible image	Manual selection + Polynomial Model	4.401	3.185	5.433	−5.133	9.701
	Grayscale and gradient information + MRF	0.712	0.630	0.951	0.888	0.791

3.3. Multimodal Image Registration of Large Deformation

In the above experiment, the initial maximum disparity of the multimodal image is about 75 pixels. A test regarding whether the framework is effective for the large parallax multimodal image or not is conducted in order to further verify the adaptability of the proposed framework. The mid-infrared image is subjected to manual translation while the mid-infrared data of the large initial disparity is experimented.

According to the maximum label value and the pyramid image series of the displacement discrete space, three experiments are set up correspondingly in this paper: the translation of the mid-infrared image in the x, y and *diagonal* direction is carried out about 75 pixels. The images will be registered to the visible orthomosaic image. The experimental configuration is similar to the above experiment.

The mid-infrared images before the registration and the checkerboard visualization results after registration are shown in Figure 8. The registration accuracy is presented in Table 4. The average target registration errors of x, y and *diagonal* direction are less than 1 pixel, while the maximum displacement errors are less than 1 pixel. Compared with the smaller deformation of the multimodal image registration, the matching accuracy of the homonymy point pairs is basically the same. Therefore, it can be assumed that the proposed framework is better to meet the registration accuracy requirements for the large deformation of the multi-modal image. Multimodal image registration experiments of large deformation illustrate that the proposed registration framework could use high-precision POS directional data and high-speed image stitching to meet the accuracy requirements of multimodal image registration.

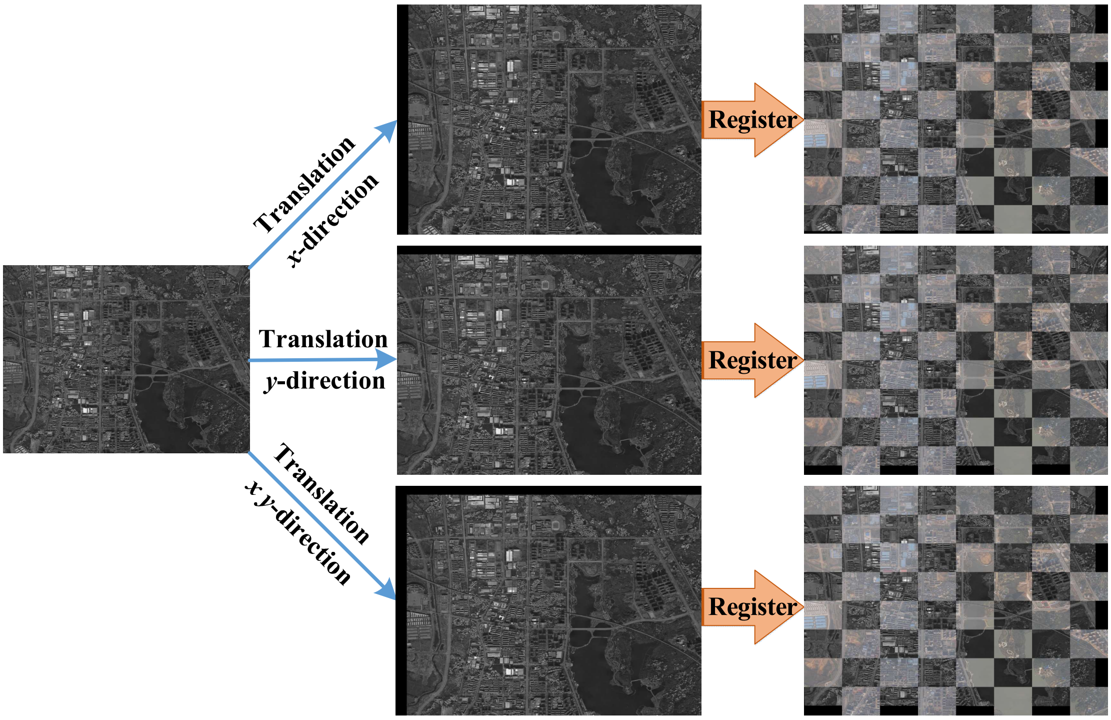


Figure 8. Mid-infrared image after translation and the registration result. The edges of the checkerboard visualization are neatly coincident.

Table 4. Quantitative evaluation result of the mid-infrared image and visible orthomosaic image after translation.

Translation Direction	*dx* (pixels)	*dy* (pixels)	*D* (pixels)	*dx_max* (pixels)	*dy_max* (pixels)
x	0.313	0.676	0.745	0.592	0.931
y	0.693	0.635	0.940	0.963	0.813
diagonal	0.148	0.517	0.538	0.258	0.956

4. Discussion

4.1. Accuracy Improvements from Gradient Information

In order to verify the performance of the framework and accuracy improvements from gradient information, these three comparative methods have been studied based on four multimodal datasets, including (1) MRF-based registration framework only using grayscale information; (2) MRF-based registration framework only using gradient information; (3) MRF-based registration framework using grayscale and gradient information.

The evaluation of three comparative registration methods was performed both qualitatively and quantitatively, the results of which were further evaluated by the local characterization of the interest region on the checkerboard visualizations for the qualitatively evaluation. After the registration, the multimodal remote sensing images were superimposed together and displayed alternately in checkerboard pattern. As can be seen in Figure 9, we selected two regions of interest on the checkerboards after three different registration methods to local zoom for the comparative analysis, which correspond to scenes with abundant textures and lack of textures, respectively. Through the observation of the checkerboard lattice edge of the multimodal image overlay, the following conclusions of the qualitative evaluation are made, which are presented below:

(1) For mid-infrared image and visible orthomosaic image registration, Worldview-3 near-infrared and multi-spectral image registration, the checkerboard visualization of the proposed framework is roughly equivalent with that of all other methods.

(2) For Terra SAR image and optical image registration, monolithic cloud point depth image and visible image registration, MRF-based registration framework only using grayscale information cannot achieve successful registration. MRF-based registration framework using grayscale and gradient information has better performance than that only using gradient information. There are still unaligned block edges on the checkerboard visualization of MRF-based registration framework only using gradient information, while checkerboard block edges of MRF-based registration framework using grayscale and gradient information are neatly coincident, so the registration effect of the proposed framework is much better than all other methods.

(3) For some scenes with abundant texture, the edge information is relatively abundant, MRF-based registration framework using grayscale and gradient information outperforms that only using grayscale information; for some scenes with lack of texture, the edge information is relatively deficient, MRF-based registration framework using grayscale and gradient information outperforms that only using gradient information. There are a large number of such areas with lack of textures in photogrammetry and remote sensing applications, so it is necessary to combine grayscale and gradient information.

In summary, the proposed framework has the best visualization results.

For quantitative evaluation, errors are in pixel level and have been calculated based on the manually-denoted homonymy points, which are all shown in Table 5. The following conclusions can be drawn:

(1) Regarding the registration accuracy, MRF-based registration framework using grayscale and gradient information has better performance than other methods. The average target registration errors of four remote sensing datasets are less than 1 pixel, while the maximum displacement errors are less than 1 pixel.

(2) The proposed framework has improved the registration accuracy of four multimodal datasets, especially the registration of Terra SAR image and optical image and the registration of monolithic cloud point depth image and visible image. The proposed framework can achieve the sub-pixel-level registration precision even in the case where other methods cannot achieve successful registration or have poor registration accuracy.

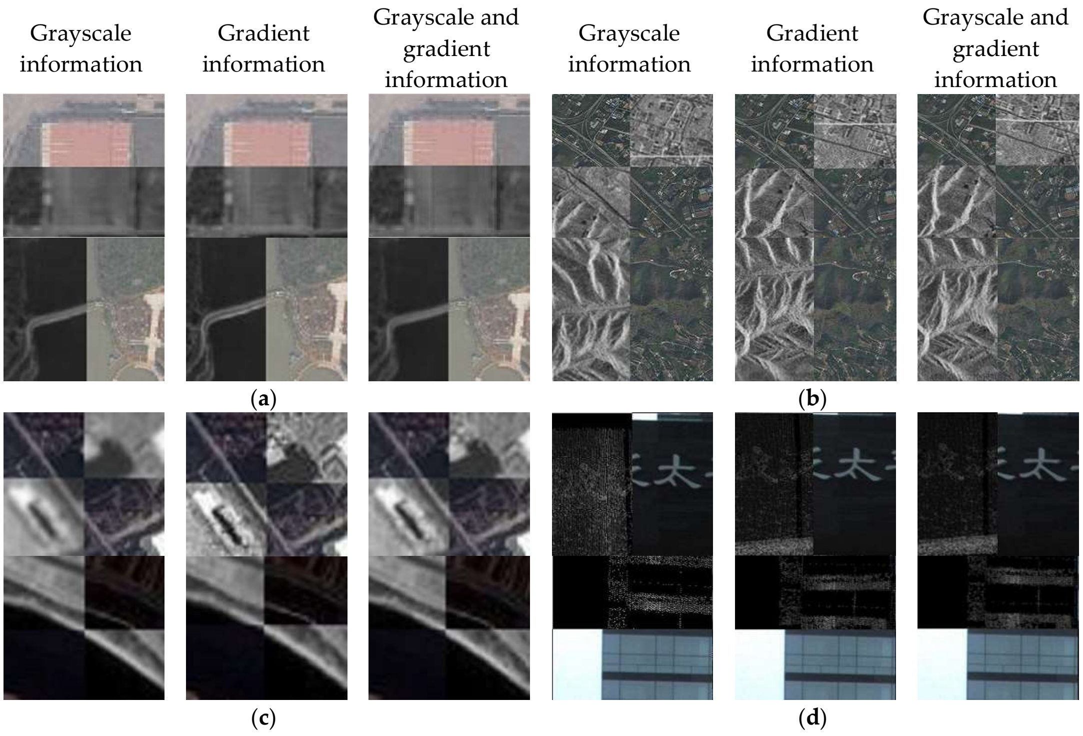


Figure 9. Local zoom graph of the interest regions on the checkerboard visualization after three comparative registration methods: (**a**) Mid-infrared image and visible orthomosaic image; (**b**) Terra SAR image and optical image; (**c**) Worldview-3 near infrared image and multispectral image; (**d**) Monolithic cloud point depth image and visible image.

Table 5. Quantitative evaluation comparison with two registration methods in four multimodal datasets.

Multimodal Dataset	Experimental Method	*dx* (pixels)	*dy* (pixels)	*D* (pixels)	*dx_max* (pixels)	*dy_max* (pixels)
Mid-infrared image and visible orthomosaic image	Grayscale information	0.435	0.797	0.908	1.374	1.822
	Gradient information	0.857	0.731	1.127	1.438	2.551
	Grayscale and gradient information	0.298	0.548	0.624	0.750	0.950
Terra SAR image and optical image	Grayscale information	8..496	9.022	12.393	12.974	11.246
	Gradient information	1.015	1.336	1.678	3.160	1.731
	Grayscale and gradient information	0.521	0.565	0.769	0.754	0.890
Worldview-3 near infrared image and multispectral image	Grayscale information	0.625	0.950	1.137	0.889	1.175
	Gradient information	0.749	0.749	1.060	1.398	1.836
	Grayscale and gradient information	0.344	0.406	0.532	0.537	0.562
Monolithic cloud point depth image and visible image	Grayscale information	1.944	5.223	5.573	5.263	7.049
	Gradient information	2.638	2.563	3.677	4.125	5.875
	Grayscale and gradient information	0.712	0.630	0.951	0.888	0.791

As shown in Figure 10, the registration error has been significantly decreased while the registration accuracy has been dramatically improved after the joint use of grayscale and gradient information.

Above all, the proposed registration framework outperforms all other methods and meets the subpixel-level registration accuracy need.

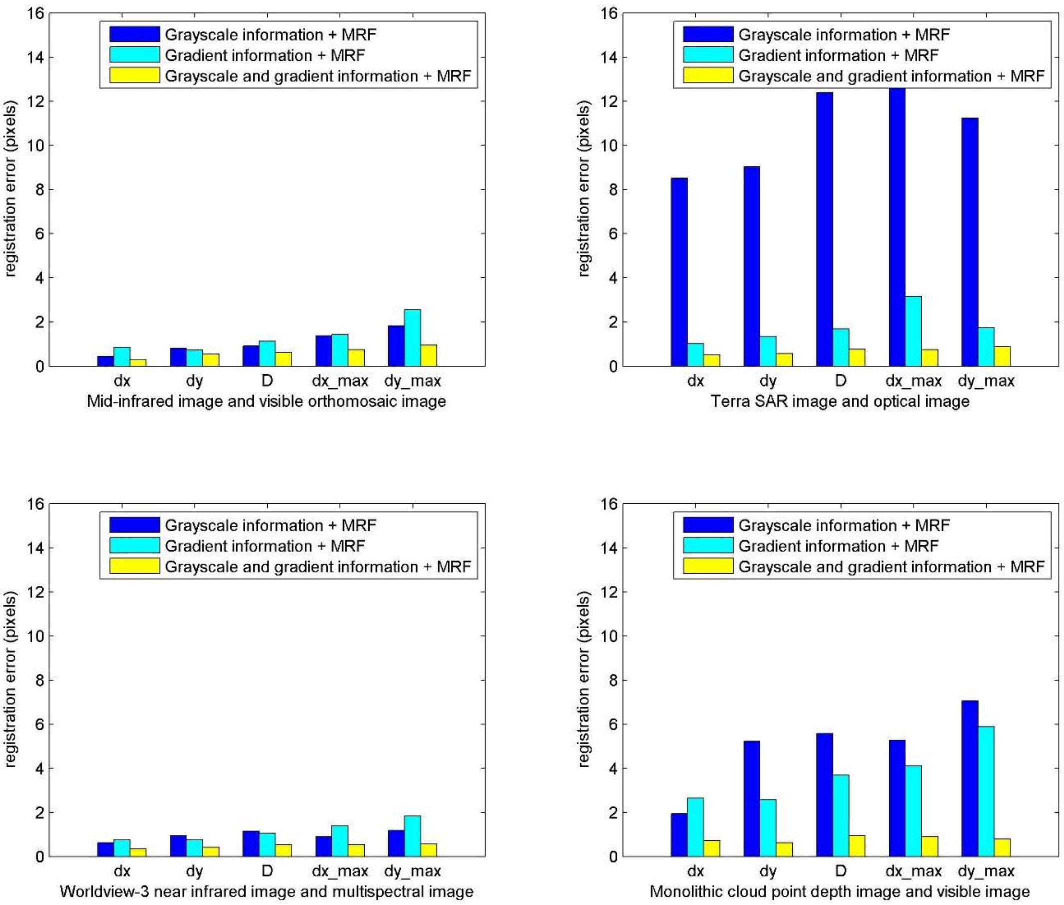


Figure 10. Accuracy improvements from gradient information in four multimodal datasets.

4.2. Accuracy Improvements from the Spatial Weighting Function η

The common spatial weight functions are listed as follows, including (1) distance threshold method; (2) distance inverse method; (3) Gaussian function method. Although the distance threshold method is simple, it is constrained by the disadvantages that the function is not continuous. Therefore, it should not be used in the registration framework. In order to investigate the accuracy improvement of the proposed framework in relation to the spatial weighting function η, we set up two experiments correspondingly: other parameters will be fixed as optimal parameter settings in each dataset while the spatial weighting function η was chosen from distance inverse method or Gaussian function method for four multimodal datasets.

As shown in Table 6, Figure 11, the use of the Gaussian function method in the proposed framework can dramatically decrease the registration error and hence to increase the registration accuracy. Above all, the proposed registration framework using Gaussian function as spatial weighting function is superior to that using distance inverse method.

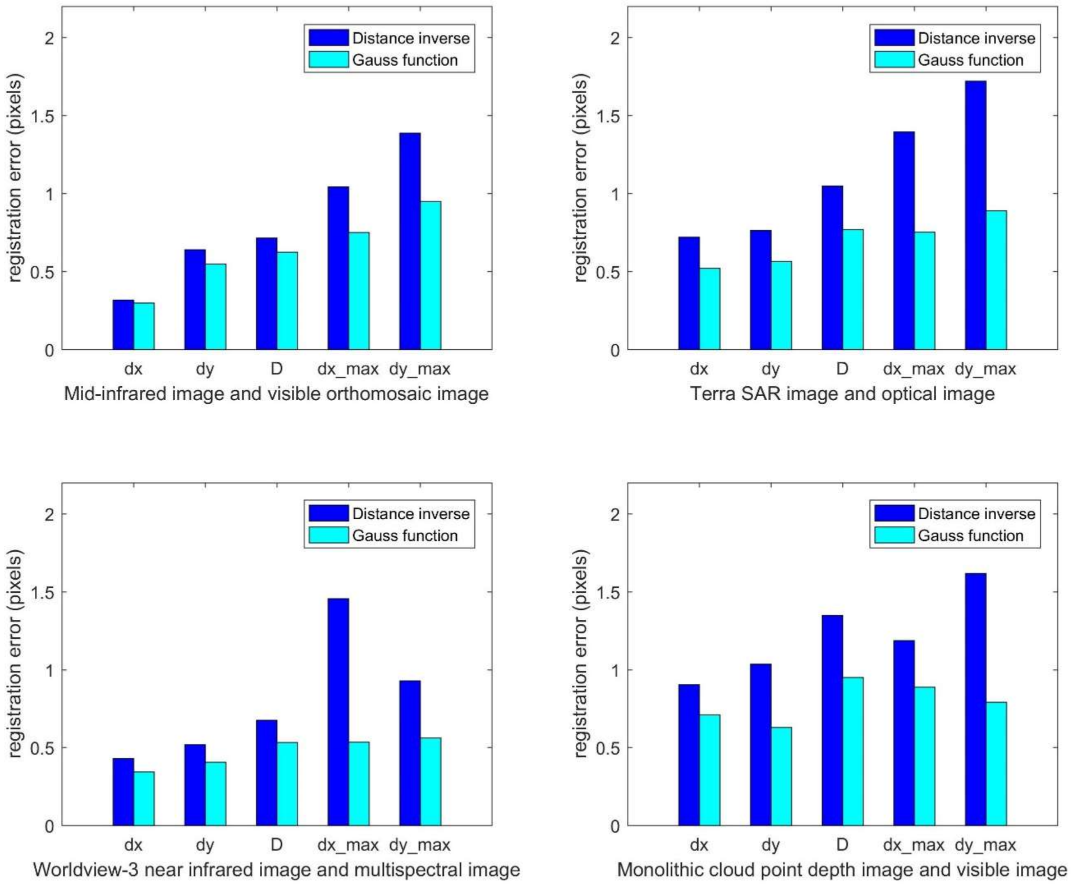


Figure 11. Accuracy improvements from spatial weighting function in four multimodal datasets.

Table 6. Quantitative evaluation comparison with the different spatial weighting function η in four multimodal datasets.

Multimodal Dataset	Spatial Weight Function	*dx* (pixels)	*dy* (pixels)	*D* (pixels)	*dx_max* (pixels)	*dy_max* (pixels)
Mid-infrared image and visible orthomosaic image	Distance inverse	0.317	0.641	0.715	1.044	1.387
	Gaussian function	0.298	0.548	0.624	0.750	0.950
Terra SAR image and optical image	Distance inverse	0.721	0.763	1.050	1.396	1.720
	Gaussian function	0.521	0.565	0.769	0.754	0.890
Worldview-3 near infrared image and multispectral image	Distance inverse	0.432	0.520	0.676	1.456	0.928
	Gaussian function	0.344	0.406	0.532	0.537	0.562
Monolithic cloud point depth image and visible image	Distance inverse	0.906	1.036	1.348	1.188	1.618
	Gaussian function	0.712	0.630	0.951	0.888	0.791

4.3. Influence of the Weight of the Regular Term λ

The weight of the regular term λ is a scalar value used to evaluate the influence of the regular term. In order to test the sensitivity of the proposed framework in relation to the weight of the regular term, the λ is varied from 200 to 2400.

As shown in Figure 12, the registration error decreases at first before a gradual decrease. The bottom of registration error reflects the most suitable weight of the regular term λ. There are four multimodal images having different suitable weights of the regular term λ. When λ is less than the most suitable weight of the regular term λ, the registration error would greatly increase while the registration results would be seemingly distorted. When λ is gradually greater than the most suitable weight of the regular term, wrong registration results would be obtained.

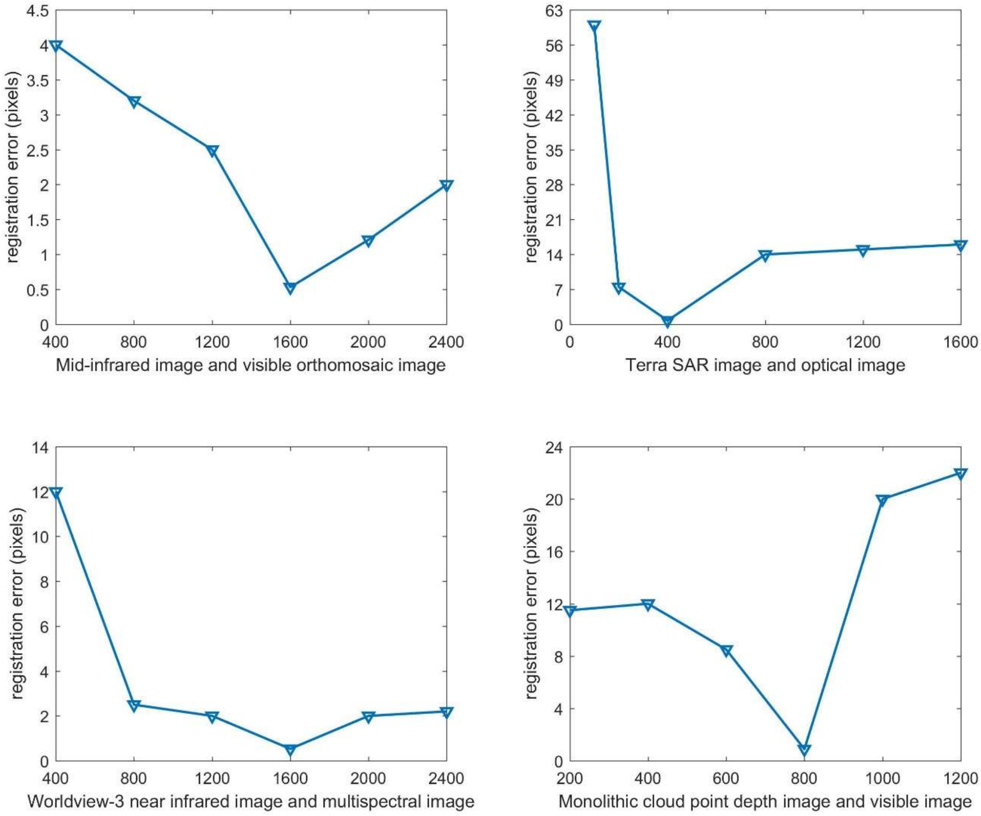


Figure 12. Comparison with the weight of the regular term λ in four multimodal datasets.

5. Conclusions

In this paper, we study the generic and automatic MRF-based registration, which can realize automatic registration with high stability and registration accuracy. Accounting for the multi-modality nature of remote sensing data, we appropriately adopt it by using the grayscale and gradient statistics information simultaneously. The edge features of the multimodal images can be well represented by the gradient information and hence to enhance the accuracy of the registration results. The spatial weighting function was optimized and has further improved the registration results. The value space was discretized to improve the convergence speed. The quantitative validation could reveal the potential of this approach on multiple multimodal remote sensing datasets. In particular, the average target registration error of the proposed framework is less than 1 pixel, while the maximum displacement error is less than 1 pixel. The proposed registration framework uses the mutual information to measure the joint probability distribution of the multimodal image, so that the initial disparity range of the image to be registered is limited. In light of this, the registration of multimodal remote sensing image with significant rotation and scale change needs be further studied. Thus, we plan to optimize the proposed approach, which will be suitable for other images with large parallax. Moreover, we plan to explore GPU implementations towards real time performances, which will be applied on large multimodal remote sensing datasets.

Author Contributions: Z.W. and Z.Y. optimized the method; Z.W. conceived and designed and performed the experiments; L.Y., Z.W., Z.Y. and Y.L. analyzed the data; Z.W. and Z.Y. evaluated the method; Z.W., L.Y. and Z.Y. wrote the paper; and Y.L. helped to prepare the manuscript. The final manuscript is the joint work of the whole team of authors.

Acknowledgments: The authors would like to show their gratitude to the editors and the anonymous reviewers for their comments and suggestions. This research was supported by National Key Research and Development Program of China under Grant 2017YFC0803802. The authors would like to thank Hong Xie from Wuhan University for providing the experimental data.

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

References

1. Liu, J.; Gong, M.; Qin, K.; Zhang, P. A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 545–559. [CrossRef] [PubMed]
2. Bongiorno, D.L.; Bryson, M.; Bridge, T.C.L.; Dansereau, D.G.; Williams, S.B. Coregistered Hyperspectral and Stereo Image Seafloor Mapping from an Autonomous Underwater Vehicle. *J. Field Robot.* **2017**. [CrossRef]
3. Sakamoto, Y.; Kuwahara, T.; Yoshida, K. HPT: A High Spatial Resolution Multispectral Sensor for Microsatellite Remote Sensing. *Sensors* **2018**, *18*, 619. [CrossRef]
4. Guo-Qiang, N.I.; Liu, Q. Analysis and prospect of multi-source image registration techniques. *Opto-Electron. Eng.* **2004**, *31*, 1–6.
5. Brown, L.G. A Survey of Image Registration Techniques. *ACM Comput. Surv.* **1999**, *24*, 325–376. [CrossRef]
6. Fan, B.; Huo, C.; Pan, C.; Kong, Q. Registration of Optical and SAR Satellite Images by Exploring the Spatial Relationship of the Improved SIFT. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 657–661. [CrossRef]
7. Li, H.; Manjunath, B.S.; Mitra, S.K. A contour-based approach to multisensor image registration. *IEEE Trans. Image Process.* **1995**, *4*, 320–334. [CrossRef] [PubMed]
8. Li, H.H.; Zhou, Y.T. Automatic visual/IR image registration. *Opt. Eng.* **1996**, *35*, 391–400. [CrossRef]
9. Zhang, Q. Automatic registration of aerophotos based on SUSAN operator. *Sci. Surv. Map.* **2006**, *32*, 60–63.
10. Kelman, A.; Sofka, M.; Stewart, C.V. Keypoint Descriptors for Matching Across Multiple Image Modalities and Non-linear Intensity Variations. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, MN, USA, 17–22 June 2007; pp. 1–7.
11. Kovesi, P. Image Features From Phase Congruency. *J. Comput. Vis. Res.* **1999**, *1*, 115–116.
12. Wong, A.; Clausi, D.A. AISIR: Automated inter-sensor/inter-band satellite image registration using robust complex wavelet feature representations. *Pattern Recogn. Lett.* **2010**, *31*, 1160–1167. [CrossRef]
13. Nan, L.; Sun, Q.; Geng, L.; Hui, L.I. An Extended SURF Descriptor and Its Application in Remote Sensing Images Registration. *Acta Geod. Cartogr. Sin.* **2013**, *42*, 383–388.
14. Collignon, A. Automated multi-modality image registration based on information theory. In *Information Processing in Medical Imaging*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995; pp. 263–274.
15. Maes, F.; Collignon, A.; Vandermeulen, D.; Marchal, G.; Suetens, P. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging* **1997**, *16*, 187–198. [CrossRef] [PubMed]
16. Viola, P.; Iii, W.M.W. Alignment by maximization of mutual information. *Int. J. Comput. Vis.* **1997**, *24*, 137–154. [CrossRef]
17. Studholme, C.; Hill, D.L.G.; Hawkes, D.J. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recogn.* **1999**, *32*, 71–86. [CrossRef]
18. Rueckert, D. Non-rigid registration using higher-order mutual information. *Proc. SPIE Int. Soc. Opt. Eng.* **2000**, *3979*, 438–447.
19. Russakoff, D.B.; Tomasi, C.; Rohlfing, T.; Maurer, C.R. Image Similarity Using Mutual Information of Regions. In Proceedings of the Computer Vision—ECCV 2004, European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004; pp. 596–607.
20. Tomaževič, D.; Likar, B.; Pernuš, F. Multi-feature mutual information image registration. *Image Anal. Stereol.* **2012**, *31*, 43–53.
21. Wang, F.; Vemuri, B.C. Non-Rigid Multi-Modal Image Registration Using Cross-Cumulative Residual Entropy. *Int. J. Comput. Vis.* **2007**, *74*, 201–215. [CrossRef] [PubMed]
22. Fan, X.; Rhody, H.; Saber, E. Automatic registration of multisensor airborne imagery. In Proceedings of the 34th Applied Imagery and Pattern Recognition Workshop (AIPR'05), Washington, DC, USA, 19 October–21 December 2005; IEEE: Piscataway, NJ, USA, 2005; Volume 12, pp. 6–86.

23. Li, Y.; Zhenling, M.A. An Operator of Gradient Consistency for Multimodal Image Registration. *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomat. Inf. Sci. Wuhan Univ.* **2013**, *38*, 969–972.
24. Heinrich, M.P.; Jenkinson, M.; Bhushan, M.; Matin, T.; Gleeson, F.V.; Brady, S.M.; Schnabel, J.A. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Med. Image Anal.* **2012**, *16*, 1423–1435. [CrossRef] [PubMed]
25. Ye, Y.; Shan, J.; Peng, J.; Xiong, J.; Li, W. Automated multispectral remote sensing image registration using local self-similarity. *Acta Geod. Cartogr. Sin.* **2014**, *43*, 268–275.
26. Pluim, J.P.W.; Maintz, J.; Viergever, M.A. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Med. Imaging* **2002**, *19*, 809–814. [CrossRef] [PubMed]
27. Guo, Y.; Lu, C.C. Multi-modality Image Registration Using Mutual Information Based on Gradient Vector Flow. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; IEEE: Piscataway, NJ, USA, 2006; Volume 3, pp. 277–284.
28. Yong, S.K.; Lee, J.H.; Ra, J.B. *Multi-Sensor Image Registration Based on Intensity and Edge Orientation Information*; Elsevier Science Inc.: New York, NY, USA, 2008; pp. 3356–3365.
29. Lee, J.H.; Yong, S.K.; Lee, D.; Kang, D.G.; Ra, J.B. Robust CCD and IR Image Registration Using Gradient-Based Statistical Information. *IEEE Signal Proc. Lett.* **2010**, *17*, 347–350.
30. Karantzalos, K.; Sotiras, A.; Paragios, N. Efficient and Automated Multimodal Satellite Data Registration through MRFs and Linear Programming. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 335–342.
31. Platias, C.; Vakalopoulou, M.; Karantzalos, K. Automatic mrf-based registration of high resolution satellite video data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *III-1*, 121–128. [CrossRef]
32. Yan, D.Q.; Liu, C.F.; Liu, S.L.; Liu, D.S. A Fast Image Registration Algorithm for Diffeomorphic Image with Large Deformation. *Acta Autom. Sin.* **2015**, *41*, 1461–1470.
33. Komodakis, N.; Tziritas, G. Approximate labeling via graph cuts based on linear programming. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1436–1453. [CrossRef] [PubMed]
34. Komodakis, N.; Tziritas, G.; Paragios, N. *Performance VS. Computational Efficiency for Optimizing Single and Dynamic MRFs: Setting the State of the Art with Primal-Dual Strategies*; Elsevier Science Inc.: New York, NY, USA, 2008; pp. 14–29.
35. Strekalovskiy, E.; Cremers, D. *Real-Time Minimization of the Piecewise Smooth Mumford-Shah Functional*; Springer: Berlin, Germany, 2014; pp. 127–141.
36. Hirschm, H. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 328–341. [CrossRef] [PubMed]
37. Kim, J.; Kolmogorov, V.; Zabih, R. Visual correspondence using energy minimization and mutual information. In Proceedings of the IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; pp. 1033–1040.
38. Choi, Y.; Lee, S. Injectivity Conditions of 2D and 3D Uniform Cubic B-Spline Functions. *Graph. Models* **2000**, *62*, 411–427. [CrossRef]
39. Li, X.; Zheng, L.; Hu, Z. SIFT based automatic registration of remotely-sensed imagery. *J. Remote Sens.* **2006**, *10*, 885–892.