Neurocomputing 392 (2020) 108–120

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multimodal image registration using histogram of oriented gradient distance and data-driven grey wolf optimizer

Xiaohu Yan[a], Yongjun Zhang[a,*], Dejun Zhang[b], Neng Hou[c]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[b] School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China
[c] School of Computer Science, Yangtze University, Jingzhou 434023, China

ARTICLE INFO

Article history:
Received 16 August 2019
Revised 6 December 2019
Accepted 29 January 2020
Available online 4 February 2020

Communicated by Bin Fan

Keywords:
Image registration
Multimodal image
Histogram of oriented gradient distance
Grey wolf optimizer
Data-driven strategy

ABSTRACT

Multimodal image registration is becoming increasingly important in remote sensing. However, due to the significant nonlinear intensity differences between multimodal images, conventional registration methods tend to get trapped into local optima. To address this issue, we present a new approach for multimodal image registration using histogram of oriented gradient distance (HOGD) and data-driven grey wolf optimizer (DDGWO). First, we propose a novel similarity measure for area-based registration methods that is HOGD. We investigate the performance of HOGD by analyzing its similarity curve. HOGD has a large range of values, which is helpful to find the global optimum. Second, we use GWO to optimize the transformation parameters. Since it is time-consuming to calculate HOGD, we propose DDGWO to minimize HOGD. In DDGWO, the iterations are divided into two parts: the training and prediction iterations. A support vector machine (SVM) regression model is trained by the historical HOGD computed in the training iterations. The trained SVM model predicts HOGD instead of calculating in the prediction iterations, which can reduce the computational time. Finally, we test the proposed approach that uses HOGD as the similarity measure and DDGWO as the search algorithm on 12 real and four simulated image pairs. Extensive experiments demonstrate that our approach saves up to 83.35-84.15% of computational time and outperforms the state-of-the-art algorithms in terms of registration accuracy.

1. Introduction

In recent years, the registration of multimodal images has become important for many applications of remote sensing, such as image fusion, image mosaic, and change detection [1,2]. The purpose of multimodal image registration is to geometrically align images of the same scene that are taken at different time, from different sensors, or from different viewpoints [3].

Image registration methods are coarsely classified into feature-based and area-based methods [4,5]. Feature-based methods first extract salient features such as edges and points, and then match them [6]. Area-based methods deal directly with the image intensity values without detecting distinct features [7]. Due to the significant nonlinear intensity differences between multimodal images, it is difficult to detect highly repeatable shared features by using feature-based methods [8]. Area-based methods which can avoid the step of feature detection are popularly and effectively used in multimodal images.

The commonly used similarity measures include the sum of squared differences (SSD), normalized cross-correlation (NCC), and mutual information (MI). SSD and NCC are sensitive to the nonlinear intensity differences between multimodal images [9]. MI overcomes the problem, and hence achieves impressive performance in multimodal image registration. Nevertheless, MI is computationally expensive, and MI-based algorithms may fail to register in complex registration cases [10].

In this study, we propose a novel similarity measure for area-based registration methods, which is histogram of oriented gradient distance (HOGD). HOG is a well-known feature descriptor, which uses the locally normalized histogram of gradient orientations features [11]. HOG has been successfully used in feature-based registration methods. Abraham et al. [12] used HOG to generate a feature vector for every detected key point, and then matched all the feature vectors. Patel et al. [13] used HOG as feature descriptor for speeded up robust features (SURF) point features to address the illumination variation between images. However, to the best of our knowledge, the distance between two

* Corresponding author.
E-mail addresses: yanxiaohu@whu.edu.cn (X. Yan), zhangyj@whu.edu.cn (Y. Zhang), zhangdejun@cug.edu.cn (D. Zhang), nhou@yangtzeu.edu.cn (N. Hou).

https://doi.org/10.1016/j.neucom.2020.01.107
0925-2312/

HOG feature vectors has not been used as a similarity measure for area-based registration methods.

To obtain the optimal similarity measure, heuristic algorithms are commonly used in area-based methods, such as genetic algorithms, simulated annealing, and particle swarm optimization (PSO) [14]. However, these search algorithms tend to get trapped into local optima because of the significant nonlinear intensity differences between multimodal images. Since grey wolf optimizer (GWO) has been shown to perform well in complex optimization problems [15,16], we apply the algorithm to multimodal image registration. Furthermore, to reduce computational time, we propose data-driven GWO (DDGWO) to minimize HOGD. This study is the first work to register multimodal remote sensing images using DDGWO.

The main contributions of this paper are as follows.

1) We propose a new similarity measure for area-based registration methods that is HOGD. We analyze the similarity curves of MI, correlation HOGD, cosine HOGD, and Euclidean HOGD. HOGD has a larger range of values compared with MI, which can help search algorithms avoid local optima.
2) GWO is used to search the optimal transformation parameters by minimizing HOGD. To reduce computational time, we propose DDGWO that combines GWO with a data-driven strategy. In DDGWO, a support vector machine (SVM) model is trained to predict HOGD instead of calculating, which can result in a significant reduction in computational time.

This paper is organized as follows. Section 2 provides an overview of image registration. In Section 3, multimodal image registration using HOGD and DDGWO is presented. In Section 4, experimental results on multimodal images are analyzed. Finally, conclusions are drawn in Section 5.

2. Related work

In this section, we briefly review two categories of image registration methods: feature-based methods and area-based methods [5,7].

2.1. Feature-based methods

In general, feature-based methods consist of three main modules: feature detection, feature description, and feature matching. Scale invariant feature transform (SIFT) and its variants are the most famous algorithms to detect features because they are invariant to scale, rotation and translation [17]. Many variants of SIFT are proposed for multimodal image registration [18]. Fan et al. [19] presented a registration algorithm for optical and synthetic aperture radar (SAR) images by exploring the spatial relationship of the improved SIFT. Li et al. [20] proposed a novel multimodal image matching based on radiation-invariant feature transform (RIFT). Lv et al. [21] presented a rapid algorithm for multimodal image registration named MM-SURF. Lv et al. [22] improved SIFT-based image registration performance by building and selecting highly discriminating descriptors. Xiang et al. [23] proposed a SIFT-like algorithm for optical-to-SAR image registration named OS-SIFT. Lv [24] presented a new registration algorithm for multimodal images named self-similarity and symmetry with SIFT (3S-SIFT). In addition, Lv et al. [25] proposed a corner based registration algorithm for multimodal images. Ye and Shen [26] presented a dense descriptor named histogram of orientated phase congruency (HOPC), which can capture structure and shape features of multimodal images.

With the help of deep learning, more and more learning based descriptors appear. Han et al. [27] proposed a unified algorithm for image matching that jointly learns a deep neural network for local patch representation as well as a network for feature comparison. Simo-Serra et al. [28] trained a Siamese network to learn discriminant patch representations. Wu et al. [29] used a convolutional-stacked autoencoder network to extract intrinsic deep features. Tian et al. [30] proposed to learn high performance descriptor which can be matched by L2 distance. Luo et al. [31] proposed a unified learning framework that leverages and aggregates the cross-modality contextual information. Yi et al. [32] presented a new deep network architecture that implements detection, orientation estimation, and feature description. Shen et al. [33] presented a new end-to-end trainable matching network based on receptive field. Ono et al. [34] proposed a novel deep architecture and a training strategy to learn a local feature pipeline from scratch.

The above feature-based methods have effectively improved registration accuracy. However, due to the significant nonlinear intensity differences between multimodal images, feature-based methods cannot detect highly repeatable common features, and hence show poor registration performance.

2.2. Area-based methods

Area-based methods can be generally classified into three categories: correlation-like methods, Fourier methods, and MI methods.

Correlation-like methods calculate the similarities of window pairs in two images, and consider the one with the largest similarity as a correspondence [5]. The methods have two drawbacks: the flatness of the similarity measure maxima and high computational complexity. However, the methods are still often in use because of their easy hardware implementation [35].

Fourier methods exploit the Fourier representation of images in the frequency domain, and search for the optimal spectral match [36]. The methods are robust to the frequency-dependent noise and non-uniform, time-varying illumination disturbances [37]. Nevertheless, Fourier methods may fail to match when there are significantly different spectral contents between images.

MI methods are commonly used in the registration of multimodal images. In MI methods, the objective is to maximize MI between two images. Chen et al. [38] investigated the use of a new joint histogram estimation algorithm named generalized partial volume estimation (GPVE) to compute MI. An et al. [39] used a modified PSO method named CRI-PSO, which reinitializes particle velocity to search the maximum MI. Gong et al. [40] presented a coarse-to-fine algorithm for image registration based on SIFT and MI. Fan et al. [41] proposed an improved MI method that combines the spatial information through a feature-based selection mechanism. Liang et al. [42] proposed a novel similarity measure based on spatial and mutual information (SMI), and adopted ant colony optimization (ACO) to optimize SMI. Wu et al. [43] combined ACO and local search to maximize MI. Despite their outstanding performance, MI methods are computationally expensive and tend to get trapped into local optima in multimodal image registration.

Recently, researchers have attempted to directly learn the geometric transformation of multimodal images. Miao et al. [44] proposed a CNN regression approach to estimate the transformation parameters of medical images. Cao et al. [45] used a CNN based regression model to directly learn the complex mapping between images. De et al. [46] proposed a CNN that analyzes an image pair and outputs parameters for the spatial transformer. Balakrishnan et al. [47] introduced an unsupervised learning-based algorithm for deformable medical image registration. Wang et al. [48] presented an end-to-end architecture that learns the mapping function between images and their matching labels. Shen et al. [49] proposed a deep-learning framework that combines an affine registration and a vector momentum-parameterized stationary velocity field model. Due to the local noise and intensity differences

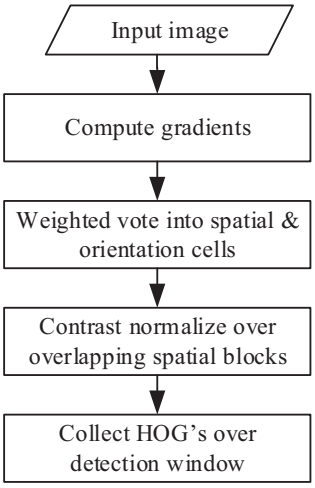


Fig. 1. Flowchart of HOG.

between multimodal images, it is difficult to train a suitable network for registration. Moreover, a large number of multimodal images are needed to train a deep network, which is challenging for multimodal remote sensing images.

3. Methodology

In multimodal image registration, we use HOGD as the similarity measure and DDGWO as the search algorithm.

3.1. Transformation model

The rigid transformation model is considered due to its wide applicability [5,14,50]. Let θ denote the rotation angle. The translations of the x-axis and y-axis are denoted as t_x and t_y, respectively. In rigid transformation model, the mapping of coordinates $p = [x\ y]^{\mathrm{T}}$ into $p' = [x'\ y']^{\mathrm{T}}$ can be formulated as

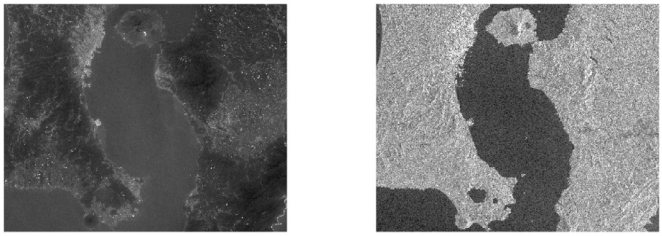
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{1}$$

Current technologies can remove obvious geometric distortions, and produce remote sensing images that have an offset of only dozen or so pixels [51]. The scale differences between remote sensing images can be removed by using their physical models [8]. Thus, the rigid transformation model is suitable for the registration of multimodal remote sensing images. The goal of search algorithms is to obtain the optimal values of t_x, t_y, and θ.

3.2. Histogram of oriented gradient distance (HOGD)

HOG is a gradient-based feature descriptor proposed by Dalal and Triggs [11]. The aim of HOG is to describe an image by capturing the statistics of oriented gradients of image pixels within the image block [52]. Due to its robustness to illumination changes and invariance to local geometric transformations, HOG has been successfully used in human detection, character recognition, and face recognition [53]. HOG performs computation on two levels of image regions: cells and blocks. Fig. 1 summarizes the flowchart of HOG.

As shown in Fig. 1, gradient orientations at each pixel are first calculated. Second, the algorithm builds a histogram of each orientation for each cell. Third, to overcome illumination variation, histograms are undergone a contrast-normalization. Finally, the combination of these histograms forms the descriptor [54].



(a) (b)

Fig. 2. A registered image pair. (a) Visible image. (b) SAR image.

In this study, we use the HOG distance of two images as the similarity measure for area-based registration methods. HOGD is robust to intensity differences between multimodal images because gradient orientations are computed from local intensity difference. Let V_r and V_s denote the HOG feature vectors of the reference and sensed images, respectively. Suppose that V_r and V_s are n-dimensional vectors. The Euclidean HOGD between two images is computed as

$$d_{euc} = \sqrt{(V_r - V_s)(V_r - V_s)'} \tag{2}$$

The cosine HOGD is computed as

$$d_{cos} = 1 - \frac{V_r V_s'}{\sqrt{(V_r V_r')(V_s V_s')}} \tag{3}$$

The correlation HOGD is computed as

$$d_{\mathrm{cor}} = 1 - \frac{(V_r - \overline{V_r})(V_s - \overline{V_s})'}{\sqrt{(V_r - \overline{V_r})(V_r - \overline{V_r})'}\sqrt{(V_s - \overline{V_s})(V_s - \overline{V_s})'}} \tag{4}$$

where $\overline{V_r} = \frac{1}{n}\sum_{j=1}^{n} V_{rj}$ and $\overline{V_s} = \frac{1}{n}\sum_{j=1}^{n} V_{sj}$. To analyze the influence of different distances, we compare the similarity curves of MI, correlation HOGD, cosine HOGD, and Euclidean HOGD. A registered pair of visible and SAR images is used to evaluate the similarity curve, which is shown in Fig. 2.

In Fig. 2, the visible image is the reference image, and the SAR image is the sensed image. When the transformation parameters are set to different values, the similarity curves are shown in Fig. 3.

In Fig. 3, the x-axis represents the translation pixel or rotation angle, and the y-axis represents the value of MI, correlation HOGD, cosine HOGD or Euclidean HOGD. The range of translation is [−10, 10], and the range of rotation is [−5, 5]. By analyzing the similarity curves in Fig. 3, the following findings are yielded:

1) The minimum of HOGD occurs at 0° of rotation or 0 pixel of translation, which confirms that the proposed similarity measure needs to be minimized to find the optimal transformation [55].
2) There are many local optima in the similarity curves of HOGD and MI. This is mainly attributed to the significant nonlinear intensity differences between multimodal images. Thus, to obtain the optimal similarity measure, it is necessary to use efficient search algorithms that have strong global search ability such as GWO and PSO.
3) The Euclidean HOGD shows significantly a larger range of values than the other similarity measures. This feature plays an important role in the optimization of transformation parameters because it can help search algorithms avoid local optima. Thus, we use Euclidean HOGD as the similarity measure in this study.

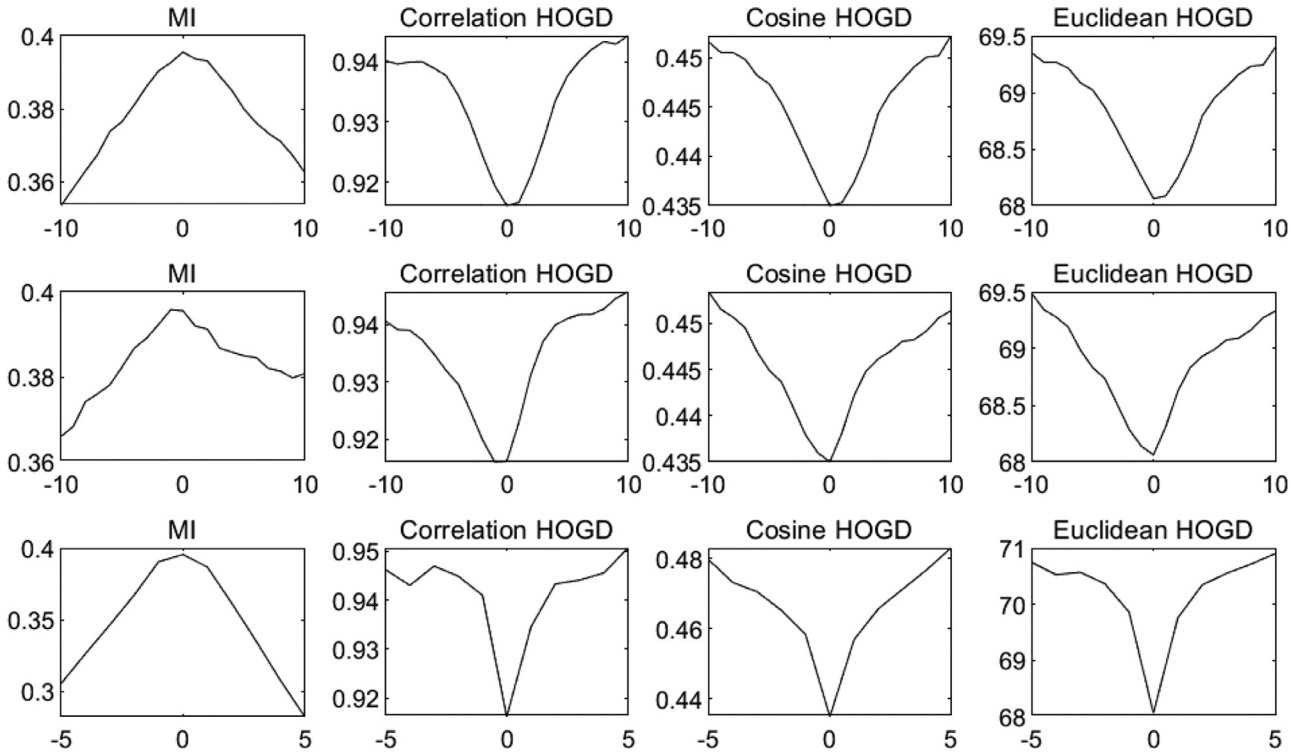


Fig. 3. Similarity curves of MI, correlation HOGD, cosine HOGD, and Euclidean HOGD. First row: Similarity curves with the *x*-axis translation. Second row: Similarity curves with the *y*-axis translation. Third row: Similarity curves with rotation.

3.3. Data-driven grey wolf optimizer (DDGWO)

GWO inspired by the hunting behavior of grey wolves is one of the latest metaheuristic algorithms. For each individual in the population, its fitness is the value of HOGD, and its position consists of the transformation parameters t_x, t_y, and θ. Then the position vector of a grey wolf can be expressed by

$$\vec{X} = (t_x, t_y, \theta) \tag{5}$$

In GWO, the distance between grey wolves and the prey is given by

$$\vec{D} = |\vec{C} \cdot \vec{X_p}(t) - \vec{X}(t)| \tag{6}$$

$$\vec{X}(t+1) = \vec{X_p}(t) - \vec{A} \cdot \vec{D} \tag{7}$$

where t indicates the iteration number, and $\vec{X_p}$ is the position vector of the prey. The coefficient vectors $\vec{A}$ and $\vec{C}$ are computed as

$$\vec{A} = 2\vec{a} \cdot \vec{r_1} - \vec{a} \tag{8}$$

$$\vec{C} = 2 \cdot \vec{r_2} \tag{9}$$

where $\vec{a}$ is linearly decreased from 2 to 0, and r_1, r_2 are random vectors in [0, 1]. In GWO, the first three best individuals obtained so far are denoted as α, β, and δ, respectively. The other wolves update their transformation parameters with respect to α, β, and δ. The mathematical equations that simulate the hunting behavior are given by

$$\vec{D_\alpha} = |\vec{C_1} \cdot \vec{X_\alpha} - \vec{X}| \tag{10}$$

$$\vec{D_\beta} = |\vec{C_2} \cdot \vec{X_\beta} - \vec{X}| \tag{11}$$

$$\vec{D_\delta} = |\vec{C_3} \cdot \vec{X_\delta} - \vec{X}| \tag{12}$$

$$\vec{X_1} = \vec{X_\alpha} - \vec{A_1} \cdot (\vec{D_\alpha}) \tag{13}$$

$$\vec{X_2} = \vec{X_\beta} - \vec{A_2} \cdot (\vec{D_\beta}) \tag{14}$$

$$\vec{X_3} = \vec{X_\delta} - \vec{A_3} \cdot (\vec{D_\delta}) \tag{15}$$

$$\vec{X}(t+1) = \frac{\vec{X_1} + \vec{X_2} + \vec{X_3}}{3} \tag{16}$$

The main difference between GWO and DDGWO lies in fitness evaluations. In DDGWO, the iterations are divided into two parts: the training and prediction iterations. An SVM regression model with Gaussian kernel is trained by the historical fitness computed in the training iterations to predict the fitness in the prediction iterations. Suppose that the percentages of the training and prediction iterations are p_1 and p_2, respectively. The pseudo code of DDGWO is presented in Algorithm 1 .

In DDGWO, the transformation parameters of each individual in the population are updated according to Eq. (16). The maximum numbers of the training and prediction iterations are $p_1 \times M$ and $p_2 \times M$, respectively. The HOGD of each individual is computed in the training iterations, while it is predicted in the prediction iterations. The trained SVM model is used to approximate fitness evaluations. Thus, the computational time can be reduced. When the termination condition is reached, the algorithm outputs the transformation parameters of the best individual $\vec{X_\alpha}$.

The percentages of the training and prediction iterations affect registration accuracy and computational efficiency. Specifically, when the percentage of the prediction iterations is too large, the computational time is reduced at the expense of registration accuracy. To balance registration accuracy and computational efficiency, we set the percentages of the training and prediction iterations to 0.15 and 0.85, respectively.

Algorithm 1: The search algorithm DDGWO.

Input: The maximum number of iterations M, the population size N, the percentage of the training iterations p_1, and the percentage of the prediction iterations p_2.
Output: The transformation parameters of $\overrightarrow{X_\alpha}$.
Initialize a, A, and C.
Randomly generate N individuals to initialize the population.
Compute the HOGD of each individual.
Compute the first three best individuals in the population that are $\overrightarrow{X_\alpha}$, $\overrightarrow{X_\beta}$, and $\overrightarrow{X_\delta}$.
for $t = 1 : p_1 \times M$ **do**
  for $i = 1 : N$ **do**
    Update the transformation parameters of the ith individual by Eq.~(16).
  end
  Update a, A, and C.
  Compute the HOGD of each individual.
  Update $\overrightarrow{X_\alpha}$, $\overrightarrow{X_\beta}$, and $\overrightarrow{X_\delta}$.
end
Train an SVM regression model using the historical HOGD computed in the training iterations.
for $t = 1 : p_2 \times M$ **do**
  for $i = 1 : N$ **do**
    Update the transformation parameters of the ith individual by Eq.~(16).
  end
  Update a, A, and C.
  Predict the HOGD of each individual by the trained SVM model.
  Update $\overrightarrow{X_\alpha}$, $\overrightarrow{X_\beta}$, and $\overrightarrow{X_\delta}$.
end

3.4. Multimodal image registration using HOGD and DDGWO

In this section, we present a registration approach for multimodal images using HOGD and DDGWO, which is named HDO. DDGWO is used to optimize the transformation parameters by minimizing HOGD. Then the flowchart of the proposed approach HDO is shown in Fig. 4.

As shown in Fig. 4, the main steps of HDO are described as follows.

1) The first step rectifies the reference and sensed images coarsely by using the direct georeferencing techniques. Hence, the obvious translation and rotation differences are removed. Moreover, the reference and sensed images are resampled to the same ground sample distance (GSD), which can eliminate scale differences.
2) The second step optimizes the transformation parameters using DDGWO. The goal of DDGWO is to obtain the optimal values of t_x, t_y, and θ by minimizing HOGD. In the training iterations, the HOGD of each individual is computed. An SVM regression model is trained by the historical HOGD that is standardized. In the prediction iterations, the trained SVM model is used to approximate HOGD by predicting instead of calculating, which can reduce the computational time. When the termination condition is satisfied, the position of the best individual $\overrightarrow{X_\alpha}$ is the best transformation parameters.
3) The third step registers the sensed image via rigid transformation. Using the best transformation parameters obtained by DDGWO, we register the sensed image according to Eq. (1).

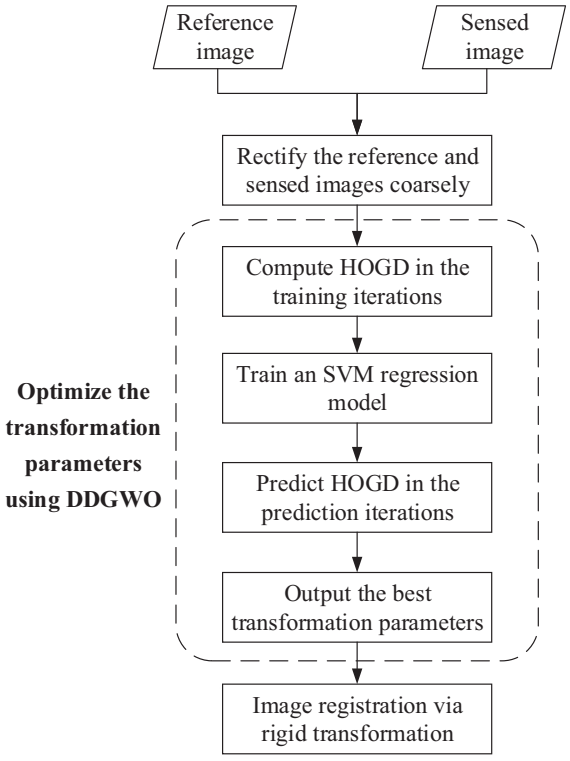


Fig. 4. Flowchart of the registration approach HDO.

3.5. Computational complexity

In HDO, the most time-consuming process is to calculate HOGD. Suppose that the time complexity of computing HOGD is $O(l)$. The maximum number of iterations is M, and the population size is N. The percentages of the training and prediction iterations are p_1 and p_2, respectively. Since HOGD is predicted by the trained SVM model in the prediction iterations, the time complexity of HDO is $O(M \times N \times l \times p_1)$. Hence, the time complexity $O(M \times N \times l \times p_2)$ can be reduced by using the proposed data-driven strategy, which results in a significant reduction in computational time.

4. Experimental results and discussions

To verify the effectiveness of the proposed approach, we compare HDO with several state-of-the-art algorithms, such as RIFT [20], HOPC [8], SIFT [17], CRI-PSO [39], and the improved SIFT (ISIFT) proposed in [22]. SIFT, ISIFT, and RIFT are used to extract feature points, and the fast sample consensus (FSC) algorithm [56] is employed to remove the outliers and to estimate the transformation parameters. CRI-PSO is an area-based method, which uses MI as the similarity measure.

The experimental analysis is structured as follows. First, we investigate the performance of HDO when its important parameters are set to different values. Second, we test HDO on multimodal images with different scale factors. Third, we analyze the registration accuracy and computational efficiency of HDO. Fourth, to evaluate the proposed similarity measure, we compare the similarity curves of HOGD and MI. Finally, we compare HDO with the registration algorithm using HOGD and GWO, which is named HGO.

4.1. Experimental setup

In area-based methods, the population size is 50, and the maximum number of iterations is 200. In feature-based methods, a

match is accepted when the distance between two images is less than three pixels. The parameters of SIFT, ISIFT, RIFT, HOPC, and CRI-PSO are set according to their original literature.

Since the obvious translation and rotation differences between the reference and sensed images are removed, the search ranges of the transformation parameters t_x, t_y, and θ are set to [−10, −10, −8; 10, 10, 8]. The algorithms are written in Matlab R2018a. All experiments are executed on an Intel(R) Core(TM) i7-8700 @3.2 GHz CPU with 8GB memory.

4.2. Evaluation criteria

The root mean square error (RMSE) and mean absolution error (MAE) of check points are used to evaluate the registration accuracy quantitatively [57]. We select check points $\{(x_i, y_i), (x_i^{'}, y_i^{'})\}$ from the reference and sensed images. Let $(x_i^{''}, y_i^{''})$ denote the transformed coordinates of $(x_i^{'}, y_i^{'})$. Then RMSE and MAE are computed by

$$RMSE = \sqrt{\frac{1}{L}\sum_{i=1}^{L}((x_i - x_i^{''})^2 + (y_i - y_i^{''})^2)} \tag{17}$$

$$MAE = \frac{1}{L}\sum_{i=1}^{L}\sqrt{((x_i - x_i^{''})^2 + (y_i - y_i^{''})^2)} \tag{18}$$

where L is the number of check points. In general, the check points are determined manually. Specifically, for each image pair, we select 40–60 evenly distributed check points with subpixel accuracy between the reference and sensed images. The runtime is used to evaluate the computational efficiency. The smaller the runtime, the higher the computational efficiency.

4.3. Description of data sets

We test HDO on the real and synthetic data sets of multimodal remote sensing images.

4.3.1. Real data sets

To evaluate the performance of HDO, the algorithm is tested on 12 pairs of real multimodal images. These images are divided into four types: 1) infrared-visible (Infra-Visib); 2) LiDAR-visible (LiDAR-Visib); 3) image-map (Img-Map); and 4) visible-SAR (Visib-SAR). Each image pair is resampled into the same GSD. Table 1 provides the descriptions of the real multimodal images, and the images are shown in Fig. 5.

The main characteristics of each image pair are summarized as follows. In Infra-Visib, image pairs 1, 2, and 3 are captured over Shandong, Liaoning, and Jiangsu Province, China, respectively. In LiDAR-Visib, the images cover urban areas with high buildings. LiDAR images have lots of noise, which makes it difficult to detect the shared feature between images. In Img-Map, image pairs 7 and 8 are captured over Qinghai Province, China; image pair 9 is captured over Tibet Province, China. In Visib-SAR, image pairs 10, 11, and 12 are captured over Shikoku, Kyushu, and Hokkaido, Japan, respectively.

As shown in Fig. 5, the real images exhibit a wide range of land covers. Moreover, the images are captured by different sensors, from different platforms, at different time, or at different bands, which can test the efficiency and robustness of the proposed approach comprehensively.

4.3.2. Synthetic data sets

To increase the registration difficulty, we test the algorithms on synthetic multimodal images. The images are simulated by the sensed images of real data sets. We choose a real image pair for each category. Specifically, we rotate the sensed images of real image pairs 1, 4, 7, and 10 by the angle 5° to produce four pairs of synthetic images. Fig. 6 shows the reference and sensed images of synthetic data sets.

Table 1
Descriptions of real data sets.

No.	Category	Image pair	Image size
1	Infra-Visib	Sentinel-2A band 8	700 × 574
		Sentinel-2A band 4	700 × 574
2		Sentinel-2B band 8	685 × 605
		Sentinel-2B band 4	685 × 605
3		Landsat 5 TM band 4	588 × 606
		Landsat 5 TM band 1	590 × 607
4	LiDAR-Visib	LiDAR height	545 × 475
		WorldView-3	545 × 475
5		LiDAR height	480 × 550
		WorldView-3	480 × 550
6		LiDAR height	524 × 524
		Ariborne visible	524 × 524
7	Img-Map	Image from Google Maps	553 × 513
		Map from Google Maps	553 × 513
8		Image from Google Maps	650 × 405
		Map from Google Maps	650 × 405
9		Image from Google Maps	660 × 508
		Map from Google Maps	655 × 502
10	Visib-SAR	Landsat 5 TM band 1	660 × 550
		Sentinel-1A	660 × 550
11		Landsat 5 TM band 1	688 × 500
		Sentinel-1A	688 × 500
12		Landsat 5 TM band 1	660 × 530
		Sentinel-1A	660 × 530

4.4. Parameter analysis

In HOGD, each block consists of $m \times m$ cells containing $h \times h$ pixels, and the adjacent blocks are overlapped. The degree of overlap is set to a half block size because a larger overlap consumes more time.

Each cell is divided into d orientation bins. The dimension of HOG depends on the number of orientation bins d. The large value of d can encode more orientation details with the high-dimensional feature vector, which is helpful to capture the shared information between multimodal images. However, the runtime increases with the number of orientation bins. Hence, we set d to 15.

The block and cell sizes affect the performance of HDO. To analyze the influence of the parameters m and h, we test HDO on the 12 real and four simulated image pairs when m and h are set to different values. The average MAE, RMSE, and runtime are presented in Table 2. In all the tables reported in this study, the best result is marked in bold.

As shown in Table 2, with the increase of the cell size, the average MAE and RMSE increase, while the average runtime decreases generally. To obtain higher registration accuracy, we set the cell size to 3 × 3. Moreover, the average MAE and RMSE are the smallest when the block size is 4. The reason is that the valuable spatial information is suppressed when the block becomes too large or too small [8]. Therefore, the cell size and the block size are set to 3 and 4, respectively.

4.5. Influence of scale difference

To investigate the influence of scale differences, we test the proposed algorithm on the real and simulated multimodal images with different scale factors. In each image pair, the scale factors of the sensed image are set to 0.5, 1.5, and 2. Table 3 displays MAE and RMSE of HDO.

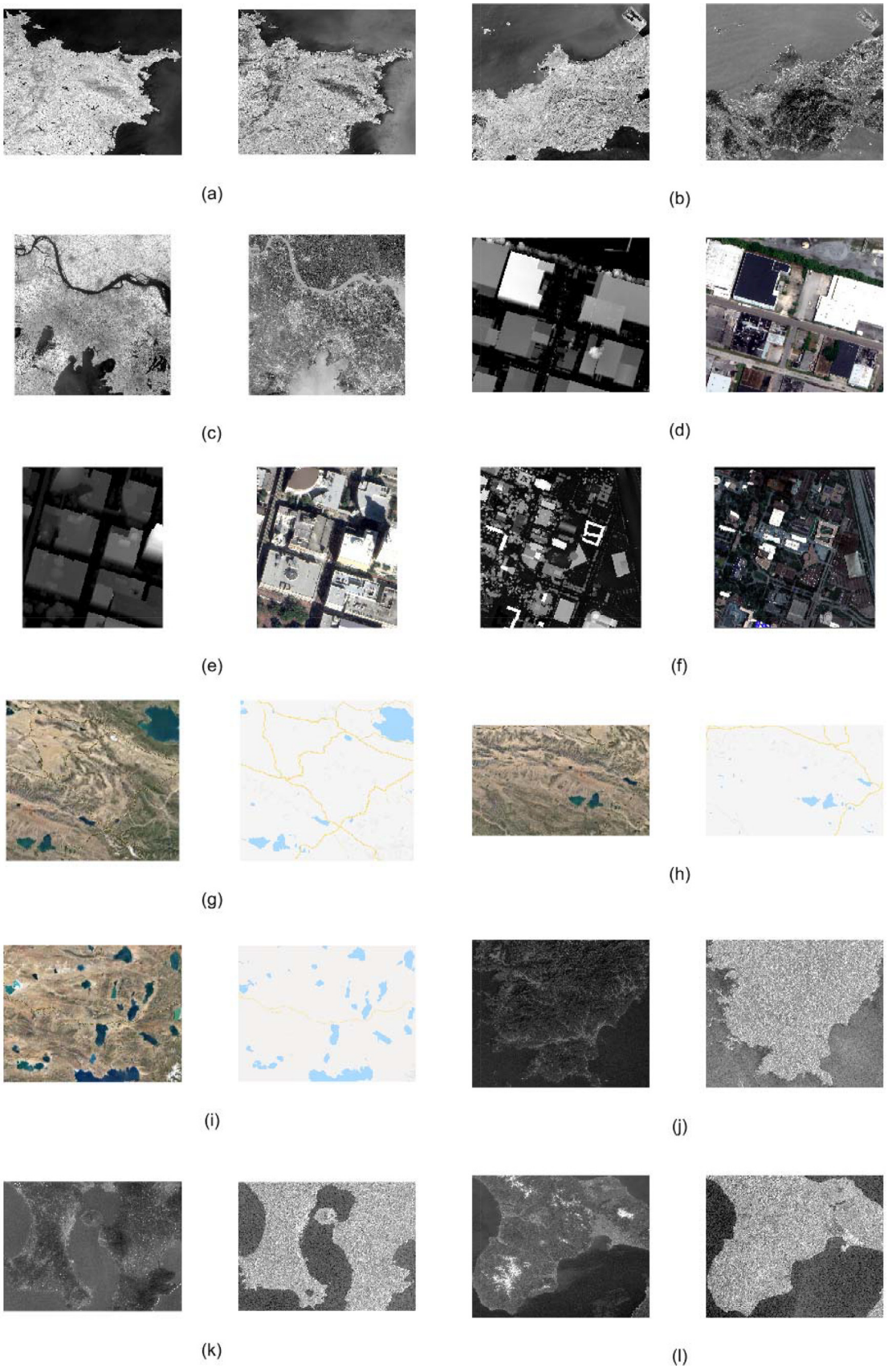


Fig. 5. Real data sets. (a) Image pair 1. (b) Image pair 2. (c) Image pair 3. (d) Image pair 4. (e) Image pair 5. (f) Image pair 6. (g) Image pair 7. (h) Image pair 8. (i) Image pair 9. (j) Image pair 10. (k) Image pair 11. (l) Image pair 12.

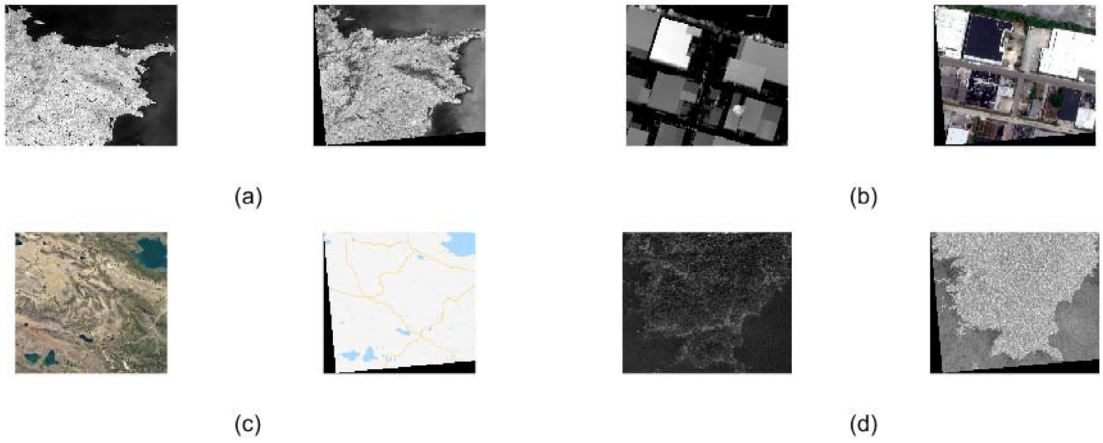


Fig. 6. Synthetic data sets. (a) Image pair 13. (b) Image pair 14. (c) Image pair 15. (d) Image pair 16.

Table 2
Average MAE, RMSE, and runtime at different block and cell sizes.

Block size	Cell size (pixels)											
	3 × 3			4 × 4			5 × 5			6 × 6		
	MAE	RMSE	Runtime	MAE	RMSE	Runtime	MAE	RMSE	Runtime	MAE	RMSE	Runtime
2 × 2	0.9798	1.2492	23.8374	1.5918	1.8044	21.0231	1.3046	1.5092	19.5975	1.1120	1.3691	**18.8561**
3 × 3	1.0193	1.2765	33.0290	1.0461	1.3240	27.1826	1.0779	1.3363	24.2717	0.8360	1.1485	22.6542
4 × 4	**0.7948**	**1.1089**	24.4931	0.9736	1.2534	21.2932	1.0710	1.3291	20.0448	1.1513	1.3634	19.2939
5 × 5	0.9307	1.2216	28.4281	1.0126	1.2932	23.9728	1.1313	1.3210	22.0739	1.0622	1.3168	21.3655
6 × 6	1.0038	1.2653	23.6805	1.1419	1.3922	20.9393	1.3627	1.6065	19.8288	1.1364	1.3778	19.0777
7 × 7	0.9513	1.2385	26.4522	1.0732	1.3009	22.7410	1.1579	1.3999	21.0586	0.9854	1.2573	20.1424

Table 3
MAE and RMSE of HDO with different scale factors.

No.	0.5		1.5		2	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	489.7924	523.8490	**1.2171**	1.3452	409.0161	437.8256
2	610.4134	617.9569	2.1609	2.3113	3.2835	3.5907
3	387.8503	431.6536	1.2385	**1.3335**	1.6989	1.8323
4	330.4742	349.1209	297.5391	314.2894	290.9078	306.1270
5	376.7029	401.3345	340.4121	362.3409	319.9520	340.5480
6	323.0442	342.6449	285.8721	303.5997	268.9280	285.6438
7	432.6381	446.5346	384.0779	396.6757	365.8796	377.5254
8	462.3339	490.3849	418.1966	443.4583	395.6708	419.2177
9	455.8561	486.9142	410.7566	438.2387	388.2028	413.9764
10	479.8616	511.6236	4.5704	5.0724	3.4511	3.6317
11	474.7201	508.1311	6.6486	7.4521	7.1512	7.8094
12	3.0118	3.1582	1.4950	1.6782	1.9925	2.1268
13	489.0663	523.1156	434.6274	465.2388	409.3403	438.1720
14	323.5727	342.4879	306.1062	322.5810	289.7941	305.0692
15	431.6340	445.5313	380.2580	392.8761	350.0087	361.4627
16	475.4910	507.8585	9.2126	9.8105	7.1551	7.5887

As can be seen from Table 3, besides image pairs 1, 2, 3, and 12, MAE and RMSE of HDO are larger than three pixels on the other image pairs. In general, MAE and RMSE of HDO are very large when there are scale differences between two multimodal images. Therefore, it is not very effective for the proposed algorithm to deal with scale differences in registering multimodal images directly. The scale differences between remote sensing images should be removed firstly by using the direct georeferencing techniques.

4.6. Analysis of registration accuracy

To analyze the registration accuracy of HDO, we compare HDO with SIFT, ISIFT, RIFT, HOPC, and CRI-PSO. MAE and RMSE of the algorithms are presented in Table 4.

As can be seen from Table 4, MAE and RMSE of HDO are smaller than those of the other algorithms on most image pairs, which confirms its higher registration accuracy. However, CRI-PSO is superior to the other algorithms on image pairs 11 and 15, and SIFT outperforms the other algorithms on image pair 13. MAE of RIFT is the smallest on image pairs 10 and 16, and RMSE of HOPC is the smallest on image pair 4. This can be attributed to the significant nonlinear intensity differences between multimodal images that make the registration complex and difficult. No algorithm can outperform all others in every registration case, which is in accord with the no-free-lunch (NFL) theorem [58]. SIFT fails to register on most image pairs, which confirms that it is indeed difficult to detect highly repeatable shared features between multimodal images using feature-based methods. The

Table 4
MAE and RMSE of SIFT, ISIFT, RIFT, HOPC, CRI-PSO, and HDO.

No.	SIFT		ISIFT		RIFT		HOPC		CRI-PSO		HDO	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	0.3946	0.5048	0.4221	0.5196	0.5005	0.5836	0.5233	0.8315	0.3990	0.4889	**0.3884**	**0.4821**
2	0.2921	0.4349	0.3301	0.4398	0.4100	0.4996	0.5572	0.8177	0.2830	0.3981	**0.2585**	**0.3985**
3	335.0658	360.8981	0.5520	0.6432	0.7562	0.8351	0.7745	0.9552	0.6678	0.7197	**0.3389**	**0.5939**
4	181.4664	195.8441	266.8214	279.9324	2.2830	2.4559	1.4544	**1.5898**	1.9560	2.1118	**0.9668**	1.7365
5	303.2039	327.6830	1.5313	1.7760	2.0638	2.2623	1.1279	1.2916	1.8499	1.9908	**0.5681**	**1.0655**
6	225.2005	238.3648	3.6838	4.0782	1.2036	1.3018	0.9885	1.1178	2.4452	2.5077	**0.4867**	**0.8916**
7	2.6818	2.9703	1.7475	1.9540	1.7881	2.1373	1.5868	1.6940	1.7977	2.0303	**0.7013**	**1.3558**
8	4.4746	4.9060	2.7563	2.9768	241.0560	298.4129	1.7445	1.8505	2.5927	2.8402	**0.4799**	**0.9298**
9	1.8258	2.2617	2.9447	3.3820	2.9852	3.3735	1.5083	1.6852	1.4714	1.8112	**0.9232**	**1.5758**
10	4.9014	5.5447	5.4844	6.1378	**0.8278**	0.9693	1.2152	1.3782	0.9100	1.0458	0.8375	**0.9561**
11	238.1188	280.3603	209.5216	214.3424	1.6939	1.8904	1.4098	1.5877	**1.0751**	**1.2894**	1.3335	1.4772
12	0.9297	1.1107	1.5860	1.7695	1.0598	1.1496	1.2355	1.3875	0.9900	1.1210	**0.5638**	**0.8307**
13	**0.4395**	**0.5434**	0.4490	0.5474	0.4743	0.5882	1.1353	1.3147	0.4748	0.5631	0.4888	0.6163
14	177.8483	187.8674	80.8960	86.5814	1.7826	2.1896	1.7092	1.8577	1.4449	1.5638	**1.3008**	**1.5152**
15	159.7547	200.8044	1.7734	2.0177	3.5280	3.8082	1.8367	1.9760	**1.2697**	**1.4590**	1.4996	1.8259
16	4.4417	4.6484	1.4193	1.5768	**1.1416**	1.2569	1.7829	1.9426	1.1489	**1.2548**	1.4550	1.5566

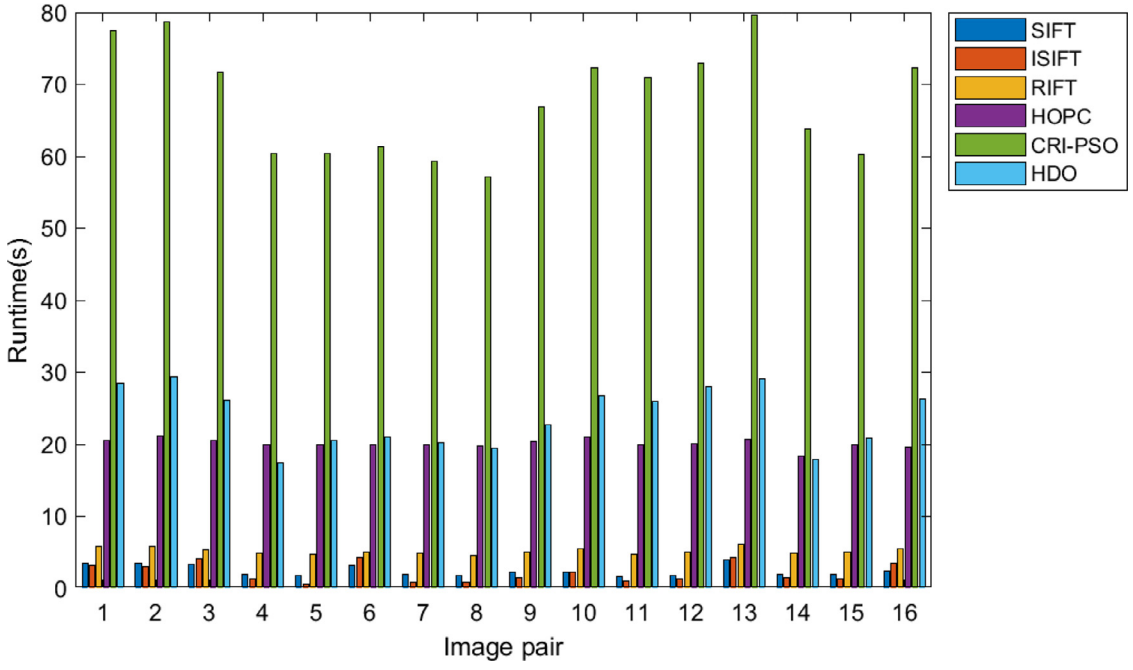


Fig. 7. Runtime comparison of SIFT, ISIFT, RIFT, HOPC, CRI-PSO, and HDO.

registration of HDO is satisfactory and accurate on all image pairs, which demonstrates the robustness and effectiveness of HDO.

4.7. Analysis of computational efficiency

The runtime is used for the evaluation of computational efficiency. We compare the runtime of SIFT, ISIFT, RIFT, HOPC, CRI-PSO, and HDO in Fig. 7. In Fig. 7, the *y*-axis represents the value of runtime in seconds.

As shown in Fig. 7, the runtime of HDO is larger than that of SIFT, ISIFT, and RIFT on all image pairs. The reason is that HDO is an area-based method, while SIFT, ISIFT, and RIFT are feature-based methods. The runtime of HDO is comparable to that of HOPC. Although the computational cost of HOGD is high, the runtime of HDO is significantly smaller than that of CRI-PSO, which partly confirms that the computational efficiency of HDO is high. This result can be explained by the fact that HOGD is predicted by the trained SVM model in the prediction iterations. The runtime of CRI-PSO is much larger than that of the other algorithms on all image pairs because MI is computationally expensive.

4.8. Evaluation of similarity measure

To evaluate the proposed similarity measure, we compare the similarity curves of HOGD and MI on multimodal images. Due to page limitation, we select a real image pair for each category to analyze the similarity measure. Specifically, the sensed images of real image pairs 1, 4, 7, and 10 are registered by the approximate ground truth transformation that is computed by the check points. We translate and rotate the registered sensed images. Then the similarity curves of HOGD and MI are presented in Fig. 8 and Fig. 9, respectively.

As shown in Fig. 8 and Fig. 9, there are lots of local optima in the similarity curves of HOGD and MI, which confirms that the optimization problem of transformation parameters is complex. Due to the significant nonlinear intensity differences between multimodal images, the maximum of MI does not occur at 0 pixel of *x*-axis translation on image pair 4. Nevertheless, the minimum of HOGD occurs at 0° of rotation or 0 pixel of translation on image pairs 1, 4, 7, and 10, which demonstrates the good robustness of HOGD. It is worth noting that HOGD shows a larger range of val-

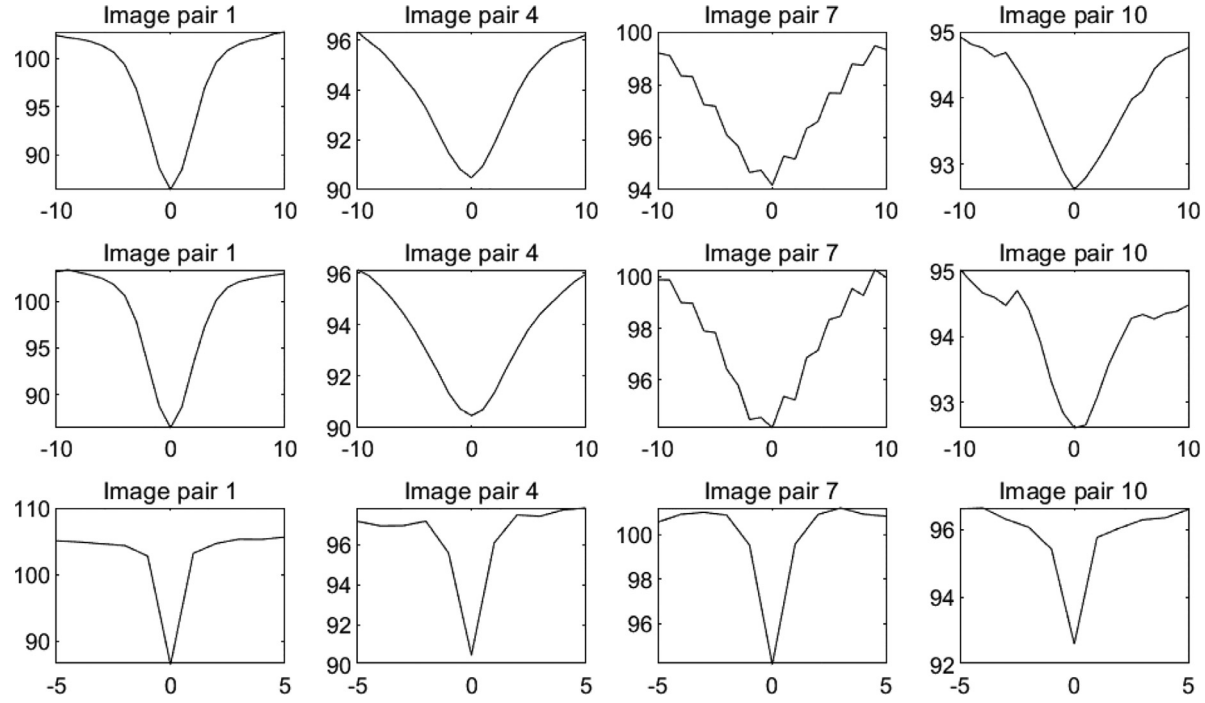


Fig. 8. Similarity curves of HOGD on image pairs 1, 4, 7, and 10. First row: Similarity curves with the *x*-axis translation. Second row: Similarity curves with the *y*-axis translation. Third row: Similarity curves with rotation.

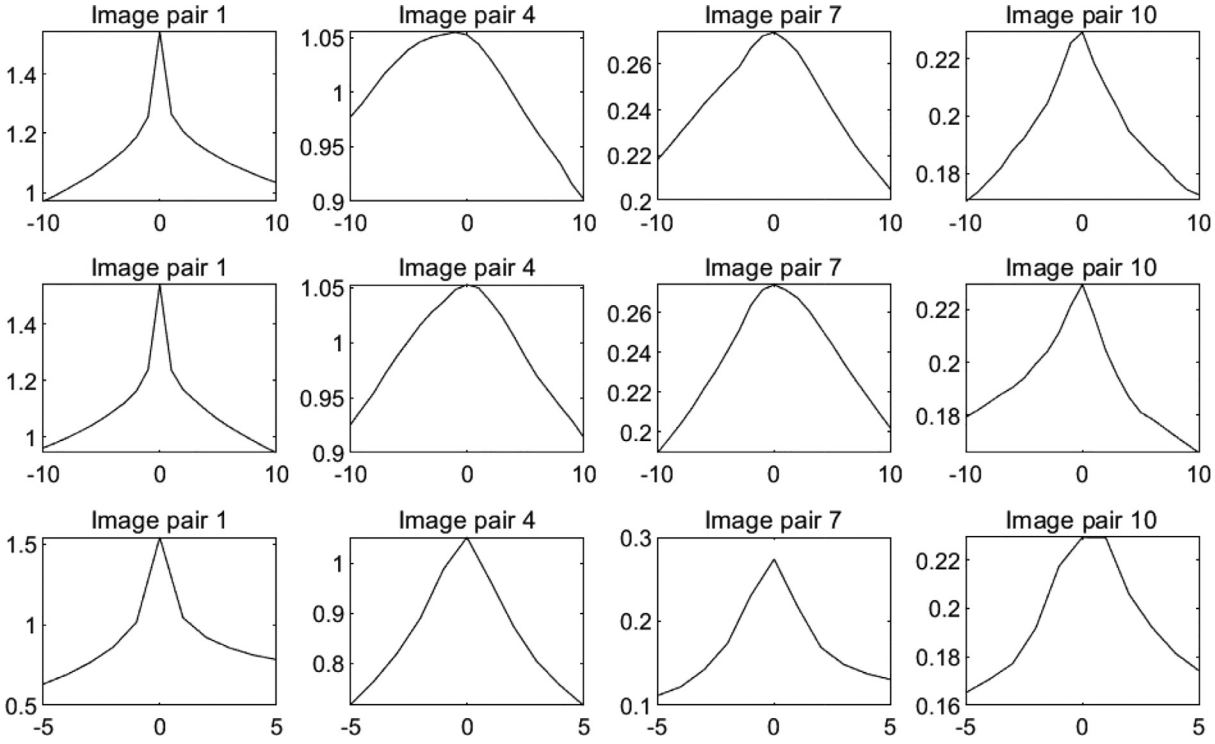


Fig. 9. Similarity curves of MI on image pairs 1, 4, 7, and 10. First row: Similarity curves with the *x*-axis translation. Second row: Similarity curves with the *y*-axis translation. Third row: Similarity curves with rotation.

ues compared with MI, which is helpful for search algorithms to find the optimal transformation parameters.

4.9. Evaluation of data-driven strategy

To investigate the influence of the proposed data-driven strategy, we compare HDO with HGO. The value of Euclidean HOGD is used to assess the search ability. To analyze the runtime conveniently, we present the improvement percentage of HDO compared to HGO. Table 5 displays the comparison results of HGO and HDO.

As shown in Table 5, compared with HGO, the runtime of HDO is significantly smaller on all image pairs. HDO saves up to 83.3537–84.1474% of computational time by using the proposed data-driven strategy. This reduction can be attributed to the fact

Table 5
Comparison results of HGO and HDO.

No.	HGO				HDO				
	MAE	RMSE	HOGD	Runtime	MAE	RMSE	HOGD	Runtime	Improvement(%)
1	**0.3686**	0.4833	**86.4233**	177.4352	0.3884	**0.4821**	86.5430	**28.2389**	84.0849
2	**0.2535**	0.4109	**90.2097**	184.7663	0.2585	**0.3985**	90.2652	**29.2903**	**84.1474**
3	**0.3587**	**0.5978**	**90.4258**	163.3609	0.3753	0.6056	90.4472	**26.1250**	84.0078
4	**0.8446**	**1.6714**	**90.4746**	106.3497	0.9668	1.7365	90.8223	**17.6841**	83.3717
5	**0.5421**	**1.0630**	**90.7913**	127.8626	0.5681	1.0655	90.8690	**20.6077**	83.8829
6	**0.4689**	**0.8869**	**89.1300**	129.9652	0.4867	0.8916	89.2304	**21.0970**	83.7672
7	**0.7108**	**1.3575**	**94.1655**	124.5473	0.7171	1.3599	94.1896	**20.2845**	83.7134
8	**0.4683**	**0.9129**	**88.9088**	117.0715	0.4799	0.9298	88.9241	**19.2069**	83.5939
9	**0.8653**	**1.5788**	**100.1726**	138.3424	0.9334	1.5822	100.2748	**22.4313**	83.7857
10	**0.8366**	**0.9549**	**92.6221**	164.8711	0.8444	0.9635	92.7131	**26.3981**	83.9887
11	**1.0539**	**1.2502**	**90.1508**	158.0624	1.3335	1.4772	90.1850	**25.3782**	83.9442
12	**0.5582**	**0.8296**	**90.4786**	160.5788	0.5638	0.8307	90.5002	**25.7462**	83.9666
13	**0.4421**	**0.5364**	**94.1354**	175.5690	0.4888	0.6163	94.5131	**28.0345**	84.0322
14	**1.1616**	1.6568	**90.8829**	105.0072	1.3008	**1.5152**	91.0343	**17.4798**	83.3537
15	**0.8581**	**1.4180**	**95.3559**	125.1926	1.4996	1.8259	95.5194	**20.3945**	83.7095
16	**1.3704**	**1.4780**	**94.8490**	163.7390	1.4550	1.5566	94.9525	**26.1810**	84.0105

that the trained SVM model predicts HOGD instead of calculating in the prediction iterations. The improvement percentage of HDO in runtime is approximately 85% which is the percentage of the prediction iterations. This result confirms that the computational cost of training the SVM model in HDO is low.

It can be seen from the results in Table 5 that MAE, RMSE, and HOGD of HDO are comparable to those of HGO on all image pairs, which confirms that the reduction in computational time is not done at the expense of registration accuracy. HGO outperforms HDO on most image pairs because the search mechanism of GWO is disturbed by the data-driven strategy partly. However, RMSE of HDO is smaller than that of HGO on image pairs 1, 2, and 14. This result may be explained by the fact that the predicted HOGD in the prediction iterations increases the population diversity and enhances the search ability of DDGWO. In general, when MAE and RMSE of an algorithm are small, HOGD of the algorithm is small. The consistency of HOGD, RMSE, and MAE demonstrates that HOGD is an effective similarity measure for multimodal image registration.

5. Conclusions

In this paper, we propose a novel approach for multimodal image registration named HDO, which uses HOGD as the similarity measure and DDGWO as the search algorithm. In DDGWO, the SVM model trained by the historical HOGD is used to predict HOGD in the prediction iterations, which reduces the computational time significantly. We compare HDO with the state-of-the-art algorithms on 12 real and four simulated image pairs. Experimental results indicate that HDO can save up to 83.35–84.15% of computational time while keeping registration accuracy high.

HDO has some advantages that make it robust and effective for multimodal image registration. An advantage of HDO is that HOGD can help search algorithms avoid local optima. This result can be explained by the fact that HOGD has a large range of values. Another advantage of HDO is that the proposed data-driven strategy can reduce the computational time significantly without much sacrifice of registration accuracy. The reason is that HOGD is predicted by the trained SVM model in the prediction iterations.

Although HDO performs well in most registration cases, the algorithm has the following limitations. First, the proposed approach strongly depends on HOG. The performance of HDO may degrade if HOG fails to capture the shared features between multimodal images. Second, since HDO is an area-based method, the runtime of HDO is still larger than that of feature-based methods such as SIFT and RIFT. Finally, it is not very effective for HDO to deal with scale differences in registering multimodal images directly. The reference and sensed images need to be rectified coarsely by using the direct georeferencing techniques.

In the future, the following directions will be investigated. First, since it is difficult to extract the highly repeatable shared features between multimodal images, we will use deep learning networks [59–62] to extract features. Second, to further reduce computational time, graphics processing unit (GPU) will be used to accelerate the calculation process of HDO. Finally, we will further improve the search ability of DDGWO to enhance the registration performance of HDO.

Declaration of Competing Interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

CRediT authorship contribution statement

Xiaohu Yan: Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Yongjun Zhang:** Data curation, Supervision, Investigation, Funding acquisition. **Dejun Zhang:** Visualization, Formal analysis. **Neng Hou:** Software, Validation.

Acknowledgments

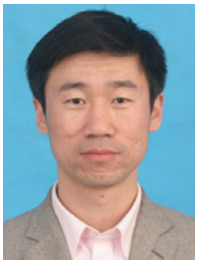
This work is supported by the National Key Research and Development Program of China under Grant 2018YFB0505003, the China Postdoctoral Science Foundation under Grant 2019M662709, and the National Natural Science Foundation of China under Grants 41871368 and 41601352.

References

[1] Y. Zhang, H. Wei, Atlas construction of cardiac fiber architecture using a multimodal registration approach, Neurocomputing 259 (2017) 219–225.
[2] Y. Ma, J. Chen, C. Chen, F. Fan, J. Ma, Infrared and visible image fusion using total variation model, Neurocomputing 202 (2016) 12–19.
[3] S. Yang, J. Zhang, W. Zhang, Phase-sensitive periodical correlation of local beam descriptors for image registration, Neurocomputing 173 (2016) 1694–1705.
[4] S. Zhao, G. Yu, A new image registration algorithm using sdtr, Neurocomputing 234 (2017) 174–184.
[5] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, J. Tian, Robust feature matching for remote sensing image registration via locally linear transforming, IEEE Trans. Geosci. Remote Sens. 53 (12) (2015) 6469–6481.
[6] Y. Wan, Y. Zhang, X. Liu, An a-contrario method of mismatch detection for two-view pushbroom satellite images, ISPRS J. Photogramm. Remote Sens. 153 (2019) 123–136.

[7] J. Li, Q. Hu, M. Ai, R. Zhong, Robust feature matching via support-line voting and affine-invariant ratios, ISPRS J. Photogramm. Remote Sens. 132 (2017) 61–76.
[8] Y. Ye, J. Shan, L. Bruzzone, L. Shen, Robust registration of multimodal remote sensing images based on structural similarity, IEEE Trans. Geosci. Remote Sens. 55 (5) (2017) 2941–2958.
[9] Y. Hel-Or, H. Hel-Or, E. David, Matching by tone mapping: photometric invariant template matching, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 317–330.
[10] Y. Keller, A. Averbuch, Multisensor image registration via implicit similarity, IEEE Trans. Pattern Anal. Mach. Intell. 28 (5) (2006) 794–801.
[11] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.
[12] E. Abraham, S. Mishra, N. Tripathi, G. Sukumaran, Hog descriptor based registration (a new image registration technique), in: 2013 15th International Conference on Advanced Computing Technologies (ICACT), 2013, pp. 1–4.
[13] M.I. Patel, V.K. Thakar, S.K. Shah, Image registration of satellite images with varying illumination level using hog descriptor based surf, Proc. Comput. Sci. 93 (2016) 382–388.
[14] M.P. Wachowiak, R. Smolikova, Y. Zheng, J.M. Zurada, A.S. Elmaghraby, An approach to multimodal biomedical image registration utilizing particle swarm optimization, IEEE Trans. Evolut. Comput. 8 (3) (2004) 289–301.
[15] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61.
[16] E. Emary, H.M. Zawbaa, A.E. Hassanien, Binary grey wolf optimization approaches for feature selection, Neurocomputing 172 (2016) 371–381.
[17] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[18] G. Lv, Robust and effective techniques for multi-modal image registration, Monash University, 2015 Ph.D. thesis.
[19] B. Fan, C. Huo, C. Pan, Q. Kong, Registration of optical and sar satellite images by exploring the spatial relationship of the improved sift, IEEE Geosci. Remote Sens. Lett. 10 (4) (2013) 657–661.
[20] J. Li, Q. Hu, M. Ai, RIFT: Multi-modal image matching based on radiation-invariant feature transform, arXiv preprint arXiv:1804.09493 (2018).
[21] D. Zhao, Y. Yang, Z. Ji, X. Hu, Rapid multimodality registration based on mm–surf, Neurocomputing 131 (2014) 87–97.
[22] G. Lv, S.W. Teng, G. Lu, Enhancing sift-based image registration performance by building and selecting highly discriminating descriptors, Pattern Recognit. Lett. 84 (2016) 156–162.
[23] Y. Xiang, F. Wang, H. You, Os-sift: a robust sift-like algorithm for high-resolution optical-to-sar image registration in suburban areas, IEEE Trans. Geosci. Remote Sens. 56 (6) (2018) 3078–3090.
[24] G. Lv, Self-similarity and symmetry with sift for multi-modal image registration, IEEE Access 7 (2019) 52202–52213.
[25] G. Lv, S.W. Teng, G. Lu, Coreg: a corner based registration technique for multimodal images, Multim. Tools Appl. 77 (10) (2018) 12607–12634.
[26] Y. Ye, L. Shen, Hopc: a novel similarity metric based on geometric structural properties for multi-modal remote sensing image matching, ISPRS Ann. Photogramm., Remote Sens. Spat. Inf. Sci. III-1 (2016) 9–16.
[27] X. Han, T. Leung, Y. Jia, R. Sukthankar, A.C. Berg, Matchnet: unifying feature and metric learning for patch-based matching, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3279–3286.
[28] E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, F. Moreno-Noguer, Discriminative learning of deep convolutional feature point descriptors, in: The IEEE International Conference on Computer Vision (ICCV), 2015, pp. 118–126.
[29] G. Wu, M. Kim, Q. Wang, B.C. Munsell, D. Shen, Scalable high-performance image registration framework by unsupervised deep feature representations learning, IEEE Trans. Biomed. Eng. 63 (7) (2016) 1505–1516.
[30] Y. Tian, B. Fan, F. Wu, L2-net: deep learning of discriminative patch descriptor in euclidean space, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 661–669.
[31] Z. Luo, T. Shen, L. Zhou, J. Zhang, Y. Yao, S. Li, T. Fang, L. Quan, Contextdesc: local descriptor augmentation with cross-modality context, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2527–2536.
[32] K.M. Yi, E. Trulls, V. Lepetit, P. Fua, Lift: learned invariant feature transform, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016, pp. 467–483.
[33] X. Shen, C. Wang, X. Li, Z. Yu, J. Li, C. Wen, M. Cheng, Z. He, Rf-net: an end-to-end image matching network based on receptive field, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 8132–8140.
[34] Y. Ono, E. Trulls, P. Fua, K.M. Yi, Lf-net: learning local features from images, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems 31, Curran Associates, Inc., 2018, pp. 6234–6244.
[35] S. Kaneko, Y. Satoh, S. Igarashi, Using selective correlation coefficient for robust image registration, Pattern Recognit. 36 (5) (2003) 1165–1173.
[36] Yun He, A.B. Hamza, H. Krim, A generalized divergence measure for robust image registration, IEEE Trans. Signal Process. 51 (5) (2003) 1211–1220.
[37] B.S. Reddy, B.N. Chatterji, An fft-based technique for translation, rotation, and scale-invariant image registration, IEEE Trans. Image Process. 5 (8) (1996) 1266–1271.
[38] H.-M. Chen, P.K. Varshney, M.K. Arora, Performance of mutual information similarity measure for registration of multitemporal remote sensing images, IEEE Trans. Geosci. Remote Sens. 41 (11) (2003) 2445–2454.
[39] R. An, P. Gong, H. Wang, X. Feng, P. Xiao, Q. Chen, Q. Zhang, C. Chen, P. Yan, A modified pso algorithm for remote sensing image template matching, Photogramm. Eng. Remote Sens. 76 (4) (2010) 379–389.
[40] M. Gong, S. Zhao, L. Jiao, D. Tian, S. Wang, A novel coarse-to-fine scheme for automatic image registration based on sift and mutual information, IEEE Trans. Geosci. Remote Sens. 52 (7) (2014) 4328–4338.
[41] X. Fan, H. Rhody, E. Saber, A spatial-feature-enhanced mmi algorithm for multimodal airborne image registration, IEEE Trans. Geosci. Remote Sens. 48 (6) (2010) 2580–2589.
[42] J. Liang, X. Liu, K. Huang, X. Li, D. Wang, X. Wang, Automatic registration of multisensor images using an integrated spatial and mutual information (smi) metric, IEEE Trans. Geosci. Remote Sens. 52 (1) (2014) 603–615.
[43] Y. Wu, W. Ma, Q. Miao, S. Wang, Multimodal continuous ant colony optimization for multisensor remote sensing image registration with local search, Swarm Evolut. Comput. 47 (2019) 89–95.
[44] S. Miao, Z.J. Wang, R. Liao, A cnn regression approach for real-time 2d/3d registration, IEEE Trans. Med. Imaging 35 (5) (2016) 1352–1363.
[45] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, D. Shen, Deformable image registration based on similarity-steered cnn regression, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 300–308.
[46] B.D. de Vos, F.F. Berendsen, M.A. Viergever, M. Staring, I. Išgum, End-to-end unsupervised deformable image registration with a convolutional neural network, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2017, pp. 204–212.
[47] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, An unsupervised learning model for deformable medical image registration, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 9252–9260.
[48] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, L. Jiao, A deep learning framework for remote sensing image registration, ISPRS J. Photogramm. Remote Sens. 145 (2018) 148–164.
[49] Z. Shen, X. Han, Z. Xu, M. Niethammer, Networks for joint affine and non-parametric image registration, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4224–4233.
[50] X. Yan, F. He, Y. Zhang, X. Xie, An optimizer ensemble algorithm and its application to image registration, Integr. Comput.-Aided Eng. 26 (4) (2019) 311–327.
[51] H. Gonçalves, J.A. Gonçalves, L. Corte-Real, A.C. Teodoro, Chair: automatic image registration based on correlation and hough transform, Int. J. Remote Sens. 33 (24) (2012) 7936–7968.
[52] S. Tian, U. Bhattacharya, S. Lu, B. Su, Q. Wang, X. Wei, Y. Lu, C.L. Tan, Multilingual scene character recognition with co-occurrence of histogram of oriented gradients, Pattern Recognit. 51 (2016) 125–134.
[53] C. Shu, X. Ding, C. Fang, Histogram of the oriented gradient for face recognition, Tsinghua Sci. Technol. 16 (2) (2011) 216–224.
[54] N. Dalal, B. Triggs, C. Schmid, Human detection using oriented histograms of flow and appearance, in: Computer Vision – ECCV 2006, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 428–441.
[55] F. Wang, B.C. Vemuri, Non-rigid multi-modal image registration using cross-cumulative residual entropy, Int. J. Comput. Vis. 74 (2) (2007) 201–215.
[56] Y. Wu, W. Ma, M. Gong, L. Su, L. Jiao, A novel point-matching algorithm based on fast sample consensus for image registration, IEEE Geosci. Remote Sens. Lett. 12 (1) (2015) 43–47.
[57] J. Li, Q. Hu, M. Ai, Lam: locality affine-invariant feature matching, ISPRS J. Photogramm. Remote Sens. 154 (2019) 28–40.
[58] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Trans. Evolut. Comput. 1 (1) (1997) 67–82.
[59] B. Ding, C. Long, L. Zhang, C. Xiao, Argan: attentive recurrent generative adversarial network for shadow detection and removal, in: The IEEE International Conference on Computer Vision (ICCV), 2019, pp. 10213–10222.
[60] K. Fu, Q. Zhao, I.Y.-H. Gu, J. Yang, Deepside: a general deep framework for salient object detection, Neurocomputing 356 (2019) 69–82.
[61] X. Kuang, X. Sui, Y. Liu, Q. Chen, G. Gu, Single infrared image enhancement using a deep convolutional neural network, Neurocomputing 332 (2019) 119–128.
[62] G. Cheng, C. Wu, Q. Huang, Y. Meng, J. Shi, J. Chen, D. Yan, Recognizing road from satellite images by structured neural network, Neurocomputing 356 (2019) 131–141.

Xiaohu Yan received the B.S. degree from Huazhong Agricultural University in 2008, the M.S. degree from North China Electric Power University in 2010, and the Ph.D. degree from Wuhan University in 2017. He is currently a Post-Doctoral Scholar in the School of Remote Sensing and Information Engineering at Wuhan University. His research interests include image registration, image matching, and optimization algorithm.



Yongjun Zhang received B.S., M.S., and Ph.D. degrees from Wuhan University, Wuhan, China, in 1997, 2000, and 2002, respectively. He is currently a Professor of photogrammetry and remote sensing in the School of Remote Sensing and Information Engineering at Wuhan University. Prof. He is the winner of the second-class National Science and Technology Progress Award (2017). He has been supported by the Changjiang Scholars Program from the Ministry of Education of China (2017), the China National Science Fund for Excellent Young Scholars (2013), and the New Century Excellent Talents in University from the Ministry of Education of China (2007). His research interests include image registration, image matching, combined block adjustment with multisource data sets, and integration of LiDAR point clouds and images.

Dejun Zhang received the Ph.D. degree from the department of computer school, Wuhan University, Wuhan, China, in 2015. He is currently a lecturer with the Faculty of Information Engineering, China University of Geosciences, Wuhan, China. Since 2015, he has been serving as a senior member of the China Society for Industrial and Applied Mathematics (CSIAM) and a committee member of the geometric design & computing of CSIAM. He is a member of the China Computer Federation (CCF). His research areas include computer graphics, computer vision, and image processing.

Neng Hou received Ph.D degree from Wuhan University in 2018. He is currently a lecture at School of Computer Science in Yangtze University. His research interests include image registration, computer vision, and GPU computing.