Neurocomputing 459 (2021) 370–382

Contents lists available at ScienceDirect

Neurocomputing

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

Fusion of multi-source retinal fundus images via automatic registration for clinical diagnosis

Tingting Dan [a], Yu Hu [a], Chu Han [b], Zhihao Fan [a], Zhuobin Huang [a], Bin Zhang [a], Guihua Tao [a], Baoyi Liu [c], Honghua Yu [c,*], Hongmin Cai [a,*]

[a] School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510006, China
[b] Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong 510080, China
[c] Department of Ophthalmology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong 510080, China

ARTICLE INFO

Article history:
Received 22 September 2020
Revised 23 May 2021
Accepted 29 May 2021
Available online 31 May 2021
Communicated by Zidong Wang

Keywords:
Retinal fundus image registration
Multiple sources
Adjustable threshold selection
Multiple features
Geometric structure constraint

ABSTRACT

Diabetic retinopathy, age-related macular degeneration and glaucoma, are the leading causes of visual impairment or blindness of the population across different ages. Retinal fundus imaging is a clinically regular tool for the diagnosis of retinal diseases. In the interest of having a comprehensive understanding of the fundus condition, it is valuable to leverage multiple fundus images from different modalities. However, a direct fusion of the multi-source fundus images eases to mis-align the physiological structure or spatial position due to possible eyeball rotations or head movements. The problem turns out to be more severe if the images were corrupted by ill conditions on eyes, such as micro-bleeding and plaques. To tackle this problem, we propose a multi-source registration model for retinal fundus images. Our proposed method considers multiple correspondences and dual structural constraints during the registration process. The method firstly selects adequate feature points by an adjustable threshold selection strategy. Then a feature-guided correspondence estimation model is established to build complementary features. Finally, their spatial transformation is built by using mean shift evolution. The evolution is guided by Tikhonov regularization on dual geometric structures. It overcomes the mess of mean shift vector field and mitigating the ill-posed displacement in field recovery. We have conducted our method on the collected 220 multi-source retinal fundus image pairs, which involve minor and larger displacement or severe retinopathy lesions, as well as additive different intensities of Gaussian noises. Extensive experiments demonstrate that the proposed method consistently outperforms seven feature-based methods.

1. Introduction

Fundus images are crucial for clinical diagnosis in ophthalmology. Retinal fundus images have always been the focus of clinicians because the pathological changes often reveal the occurrence of systematic diseases such as hypertension, diabetes, certain blood diseases, and central nervous system diseases [1]. And the deep microvascular can be observed non-invasively and directly through the retina. Therefore, an effective quantitative analysis tool of retinal images can guide the following clinical decisions and applications such as early detection, diagnosis and treatment of diseases. Currently, various types of equipment have been developed to generate different sources/modalities retinal images with different highlighted tissues, such as optical coherence tomography (OCT), color fundus photography (CFP) and fluorescent angiography (FA) [2]. For example shown in Fig. 1, the blood vessels in Fig. 1 (A) demonstrate higher intensity than the background tissue, while the blood vessels in Fig. 1 (B) are darker. It is worth obtaining comprehensive spatial information and exploit various sources/-modalities knowledge on the same tissue/region images. Broadly, image registration is a vital step for multi-source image processing (e.g., fusion). However, ill-conditioned imaging, such as misalignment, rotation or even blurry effects, may be introduced due to the eyeball or head movements. Such drawbacks lay great obstacles for accurate image registration, hindered its clinical diagnosis.

Image registration aims to align two or more images captured by multiple viewpoints, modalities or times. Conventional approaches [3,4] leverage the locally similar anchor pairs and optimize the alignment process. Feature-based approaches [5] extract various descriptive features on the images and build a transformation function to minimize the feature similarity among different modalities/times/viewpoints. However, existing methods perform unsatisfactory results in retinal image registration for ill-

* Corresponding authors.
E-mail addresses: yuhonghua@gdph.org.cn (H. Yu), hmcai@scut.edu.cn (H. Cai).

https://doi.org/10.1016/j.neucom.2021.05.091
0925-2312/

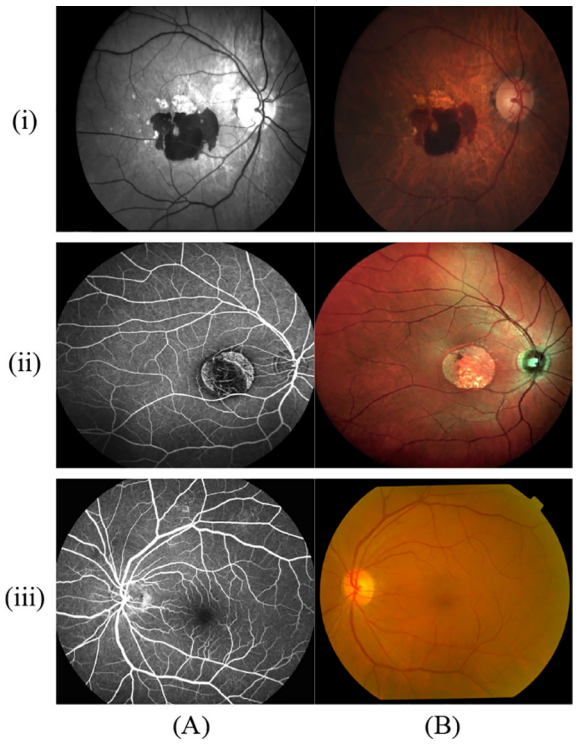

Fig. 1. Three pairs of multi-source retinal fundus images with different severity of eye conditions. (A) and (B) are collected from FA and CFP, respectively.

conditioned samples, such as severer spatial deformations, or appearing of micro-bleeding and plaques [6]. Recently, deep learning has demonstrated its merits of powerful feature representation [7,8]. Although deep learning approaches achieve outstanding performance, learning-based models require large among of training data.

Multi-source retinal image registration still suffers from the following challenges due to nonuniform contrast/intensity distributions, large homogeneous nonvascular/textureless regions, various pathology-caused degradations, and limited overlaps with few feature correspondences/matches. To address these challenges, we propose a multi-source retinal fundus image registration with feature-guided and dual structural constraints. The proposed method considers deformation recovery as a problem of clustering the kernel center to the target point set of arbitrary shapes. The major contributions of this paper are threefold:

- an adjustable threshold selection strategy is applied in feature points extraction to select the inliers for maximizing the usage of feature points;
- a multiple correspondences estimation model is established to form complementary features for enhancing the recognition of feature points;
- a spatial transformation is built by mean shift evolution, which is guided by the Tikhonov regularization on dual geometric structures. The transformation is shown to be capable of overcoming the mess of mean shift vector field and mitigating the ill-posed problem of displacement field recovery.

The rest of the paper is organized as follows. Related works are reviewed in Section 2. The method proposed to achieve retinal fundus fusion via automatic image registration is elaborated in Section 3. Extensive experiments are conducted in Section 4 to evaluate the effectiveness and robustness of the proposed method. Section 5 summarizes the discussion and conclusion of this paper.

2. Related work

The popular registration methods for retinal fundus images can be divided into two categorizes: (1) learning-based registration methods; and (2) conventional registration methods [9].

2.1. Learning-based registration method

There are few works using deep learning to handle image registration for retinal fundus images, even learning-based methods have achieved powerful performance in image classification and segmentation [10]. Mahapatra et al. [11] introduced an end-to-end deep learning method regarding retinal fundus images. The method utilizes the generative adversarial networks to finish the process of registration. Zou et al. [12] proposed an unsupervised structure-driven regression learning method. This approach first describes the complex mapping as a parameterized deformation function, and then calculates multi-scale similarity in combination with the contextual structures. These methods adopt the synthetic data or image augmentation method to augment the training data set, which are problem specific, and may lack robustness to multi-site or multi-source images. Recently, Lee et al. [13] presented a feature-based learning method for multi-modality retinal fundus images. This method first learns a deep representation that is built on a convolutional neural network, and then employs the conventional approach to complete the registration process. Wang et al. [14] introduced a content-adaptive weakly-supervised deep learning framework, which integrates the strategies of vessel segmentation, feature detection and outlier rejection. Although the deep features can learn more high-level features, the understanding of them may be inadequate and uncontrollable.

2.2. Conventional registration method

Most conventional image registration methods are generally divided into two categories: area-based methods and feature-based methods. The latter is not affected by intensity and rotation, and has less computational complexity and higher efficiency. Therefore, we focus on feature-based registration methods, which generally consist of three essential steps. The first step focuses on extracting a sufficient number of feature points from the reference image and the sensed image. The popular image descriptors, including scale-invariant feature transform (SIFT) [15], edge oriented histogram-scale invariant feature transform (EOH-SIFT) [16], and speeded up robust features (SURF) [17] are widely applied in retinal image registration [18]. The second step aims to align the feature point sets from different sources, times and viewpoints, thereby achieving correspondence estimation and transformation updating [9]. There are many different methods developed for it. Myronenko and Song [19] introduced a probabilistic method, called by coherent point drift (CPD) algorithm, which utilizes Euclidean distance to evaluate the correspondence between point sets and applies an additional uniform distribution for outlier modeling. Yang et al. [20] exploited both global and local mixture distance (GLMDTPS) features to improve the feature description for point set registration, and then proposed a multi-feature-based finite mixture model on combining SIFT with different types of geometrical features [21]. Recently, Ma et al. [22] proposed a non-rigid point set registration method by preserving global and local structures (PR-GLS) based on CPD. It first employs shape context [23] for feature point sets correspondence estimation and then assigns a priori probability manually according to the correspondence for the guide of solving the posterior probability function of the Gaussian mixture model. Subsequently, the authors applied this method to retinal image registration [24,25].

Wang et al. [26] introduced a robust multimodal retinal image registration framework (called SURF-PIIFD-RPM), which is quite robust to outliers by using the partial intensity invariant feature descriptor. Similar to SURF-PIIFD-RPM, an adaptive mismatches removing registration method (called URSIFT-RIIFD-AGMM) is proposed [27]. Its superiority can mostly be attributed to the robust initial point matching and matching postprocessing. After that, Bi et al. [18] proposed a multiple image features-based retinal image registration method (called MIF-RIRM in short), which uses multiple image features to estimate the correspondence of feature point sets, and then the global and local geometric structure constraints to control the transformation updating step. The last step is image registration/transformation, which aims at aligning the sensed image onto the reference image using the backward approach [21], and obtain the transformed image. These conventional ways increase the interpretability of the method, which are easy to understand, and have the robustness to multi-source data, as well as do not need a large number of samples for training.

3. Method

This section describes the proposed method for retinal fundus image registration. A feature point extraction scheme is firstly introduced to yield quite reliable inliers. Then a multiple feature extraction scheme is introduced to characterize each feature point comprehensively. Accordingly, we consider the deformation recovery as a problem of matching the kernel center to a target point set with arbitrary shapes. Therefore, a feature point matching method is proposed to guide the matching deformation, regulated by global and local geometric structures. The backward approach [21] is utilized to obtain the transformation for image registration. Finally, the implementation details of the proposed method are provided. The flowchart of the proposed method is illustrated in Fig. 2.

Specifically, the notations used throughout this paper is summarized in Table 1 for ease of explanation.

Table 1
List of notations used throughout the paper.

Notation	Remark
$\mathbf{I}^{\mathbf{R}}$, $\mathbf{I}^{\mathbf{S}}$, $\mathbf{I}^{\mathbf{t}'}$	the reference image, sensed image and transformed image
$\mathbf{S}=\{\mathbf{s}_n\}_{n=1}^{N}$, $\mathbf{T}=\{\mathbf{t}_m\}_{m=1}^{M}$	the feature point sets extracted from $\mathbf{I}^{\mathbf{R}}$ and $\mathbf{I}^{\mathbf{S}}$
M, N	the number of points in $\mathbf{T}$ and $\mathbf{S}$, satisfied $N \leqslant M$
$\mathbf{s}_n$, $\mathbf{t}_m$	the n-th and m-th points in $\mathbf{S}$ and $\mathbf{T}$
φ_0	the lowest threshold for extracting a large number of SIFT feature candidates in $\mathbf{I}^{\mathbf{R}}$ and $\mathbf{I}^{\mathbf{S}}$
φ^*	the highest threshold for extracting more reliable inliers
φ	the adjustable threshold changed from φ_0 to φ^* via a step size ℓ during registration
$\mathcal{T}$	the recovered transformation in every *iter* iteration
$\mathbf{CE}$	the correspondence probability matrix between $\mathbf{S}$ and $\mathbf{T}$
$\mathbf{1}$	the column vector with all ones
$\mathbf{I}$	the identity matrix
$\mathbf{0}$	the zero matrix

3.1. Feature point extraction

The emphasis in processing multi-source fundus images is to overcome the interference of a large number of outliers caused by non-rigid distortions (such as lesions) and rotations (i.e., low-overlap images). SIFT algorithm [15] that achieved excellent accuracy is employed to extract the feature points with scale and intensity invariance. However, the standard SIFT algorithm performs a feature matching before having the candidate feature points. The matching abandons a large number of feature points. To fully utilize all the feature points, an adjustable threshold selection strategy is employed. At the beginning of the iteration, a low threshold of φ_0 is used for a wider range of feature candidates. In the following iterations, the threshold will gradually increase by $\varphi = \varphi + \ell$ to win more reliable inliers until it reaches a reliable threshold of φ^*. By this strategy, more feature candidates are allowed to involve and contribute to the feature matching. The inliers can decide the overall transformation and the relevant outliers can be utilized to optimize the registration accuracy. Fig. 3 illustrates the comparison between fixed and adjustable threshold registration results.

Herein, the coordinates of extracted points are recorded as $\mathbf{P}=\{\mathbf{p}_z\}_{z=1}^{Z}$ from every image, where z is the number of extracted points. Accordingly, the source point set $\mathbf{S}=\{\mathbf{s}_n\}_{n=1}^{N}$ and the target point set $\mathbf{T}=\{\mathbf{t}_m\}_{m=1}^{M}$ are extracted from the sensed and reference images acquired by different imaging modalities, respectively.

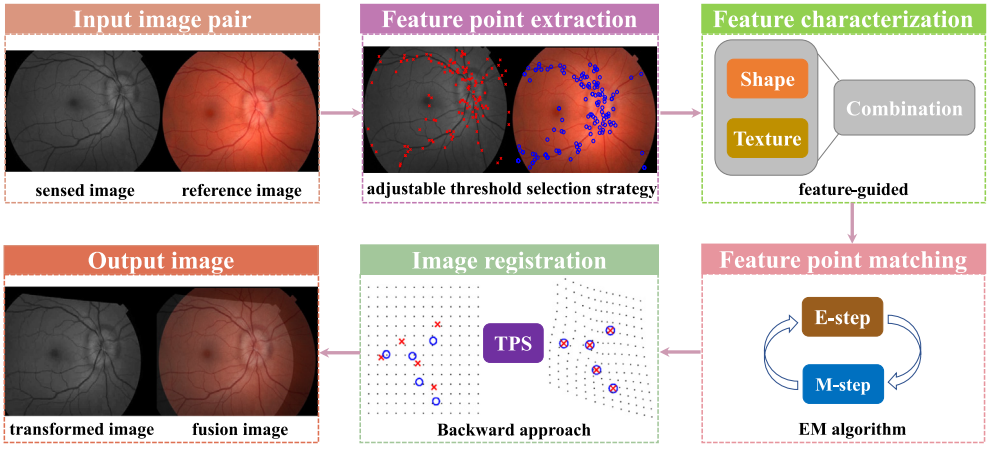

Fig. 2. The flowchart of the fully automatic multi-source retinal fundus image registration.

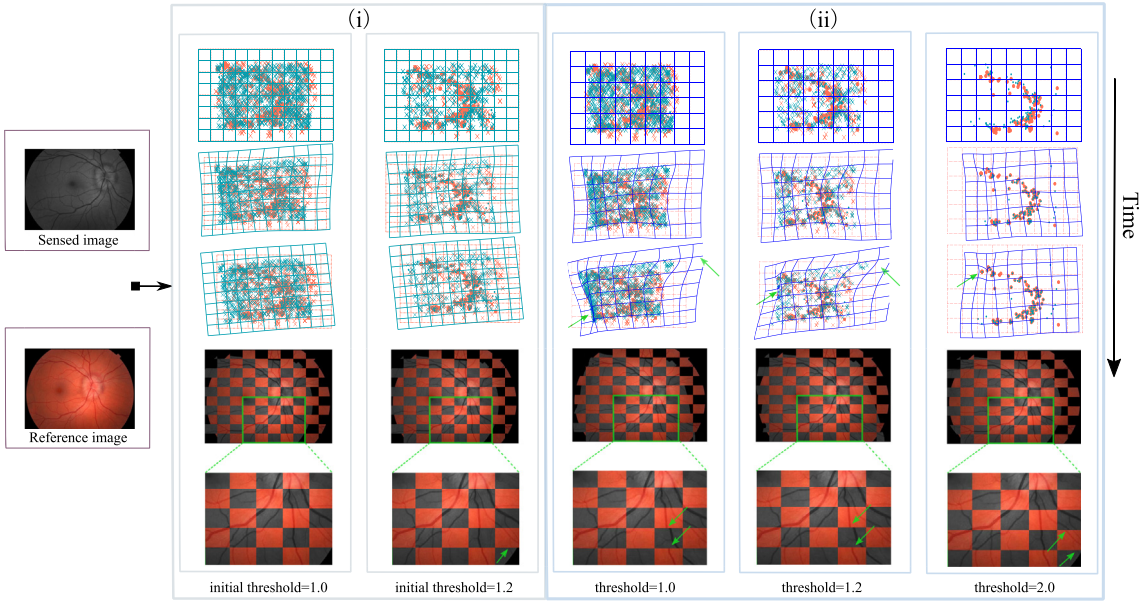

Fig. 3. Comparison results of the fixed (i) and adjustable (ii) threshold strategy on registration performance. From the second to the sixth columns, each block from top to bottom successively demonstrates: the initial pose of two feature point sets, the middle of feature point set registration, the final result of feature point set registration, the final image registration results and the enlarged exhibits of green rectangles. The orange and blue dots (crosses) denote the inliers (outliers) in two feature point sets, respectively. The orange grids (orange dotted lines) and the blue grids (blue solid lines) denote the original image field and the warped image field, respectively. For visual comparison, the image registration result (i.e., the transformed image) for each scenario is shown with the reference image by a 10×10 checkboard, where the registration errors are highlighted using the green arrows.

3.2. Feature characterization

Given the feature candidates extracted from the reference (target) and sensed (source) images, we measure their similarities by comparing both shape and textural features.

3.2.1. Measuring the matchness by shape context feature

For the shape similarity measurement, we apply shape context (SC) [23] to characterize the shape of each feature candidate. SC is originally designed as a shape descriptor which describes the spatial distribution of a shape in the log-polar diagram. It firstly sets up a polar coordinate system centered at each point and then constructs *Ra* concentric circles along the radial direction such that all the circles share *Ta* bins in the tangential direction. Therefore, the entire polar coordinate system has $B = Ra \times Ta$ bins. The number of points falling into different bins is counted and forms a histogram $\{\mathbf{h}(r,t)\}_{r=1,t=1}^{Ra,Ta}$. Finally, a χ^2-distribution is utilized to measure the difference between the source point $\mathbf{s}_n$ and the target point $\mathbf{t}_m$, denoted by a matrix **SC**.

$$\mathbf{SC}(m,n) = \frac{1}{2}\sum_{r=1}^{Ra}\sum_{t=1}^{Ta}\frac{[\mathbf{h}_{\mathbf{s}_n}(r,t) - \mathbf{h}_{\mathbf{t}_m}(r,t)]^2}{\mathbf{h}_{\mathbf{s}_n}(r,t) + \mathbf{h}_{\mathbf{t}_m}(r,t)} \tag{1}$$

Since the original design of SC only considers the points drop into each specific bin. It is sensitive to the local structure deformations introduced by the movements of organs or tissues. To alleviate such a problem, we conduct shape context descriptor by considering not only the points in the current bin but also in the neighbor ones, with an elliptic Gaussian soft counting strategy [28]. Fig. 4 visualizes the detail of this strategy. Fig. 4 (i) is the original SC and Fig. 4 (ii)–(iv) are three different counting strategies. As we can see, the neighbor bins also contribute to the central bin and the intensity of each bin indicates the weight of the contribution. This counting strategy can increase the tolerance of the deformation.

Let $\mathbf{h}(r,t)$ be the central bin, which is the integer statistical value obtained by the counting strategy of the original shape context. Let ρ_r and ρ_t be the range of radial and tangential directions, respectively. The contribution of each neighbor bin is controlled by a two-dimensional elliptic Gaussian. The re-calculated central bin $\hat{\mathbf{h}}(r,t)$ can be obtained by the following equation:

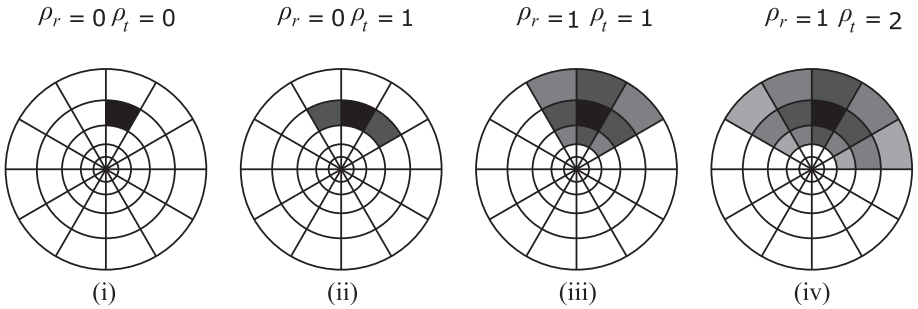

Fig. 4. Four different counting strategies. (i) is the central bin with coordinate (r,t), which is the original shape context. (ii), (iii) and (iv) demonstrate the elliptical Gaussian soft counting strategies. The intensity of the each bin indicates the weight of contribution.

$$\hat{\mathbf{h}}(r,t)=\sum_{i=-\rho_r}^{\rho_r}\sum_{j=-\rho_t}^{\rho_t}\mathbf{h}(r+i,t+j)\cdot\exp\left(-\left[\frac{i^2}{(2\rho_r+1)^2}+\frac{j^2}{(2\rho_t+1)^2}\right]\right) \tag{2}$$

The shape similarity measurement of Eq. (2) is rewritten as follows:

$$\mathbf{SC}_{mn}=\left\|\hat{\mathbf{h}}_{\mathbf{s}_n}-\hat{\mathbf{h}}_{\mathbf{t}_m}\right\|^2 \tag{3}$$

3.2.2. Measuring the matchness by texture feature

Texture feature has been widely used in the detection and recognition tasks of fundus images [29,30]. The dominant rotated local binary pattern [31] is thereby selected to represent the texture feature. For a gray-scale retinal image $\mathbf{I}_{X\times Y}$, the binary pattern feature describes the statistics of the difference distribution for gray values between the center point and the neighborhood points. It is defined as $\mathbf{BP}_{xy}=\sum_{l=0}^{L-1}2^{mod(l-D,L)}g(i(x,y),i(a_l(x),b_l(y)))$, where $(a_l(x),b_l(y))$ denotes the coordinate of the l^{th} point around the center point of (x,y), and are defined as $a_l(x)=x+Rcos(2\pi l/L)$ and $b_l(y)=y-Rsin(2\pi l/L)$, D is called the dominant direction and defined as the index of the neighboring pixel whose difference from the central pixel is maximum, i.e., $D=\arg\max_{l\in\{0,1,\ldots,L-1\}}|i(a_l(x),b_l(y))-i(x,y)|$, $i(\cdot)$ is gray value of a coordinate, R is the radius of the circle, L is the number of neighbors on the circumference, $mod(\cdot)$ represents the modulo operation. The function $g(\cdot)$ defines the binary attributes of neighborhood points, if $i(a_l(x),b_l(y))\geqslant i(x,y)$ then $g(\cdot)=1$, otherwise 0. The computed local binary pattern describes the rotation-invariant texture feature since the value of the weight term $2^{mod(l-D,L)}$ is solely determined by D. To obtain stable features, the original image is firstly weighted before extracting the texture feature $\mathbf{I}^z(x,y)=\varepsilon_{xy}^z\times\mathbf{I}(x,y)$, where $\varepsilon_{xy}^z=\exp(-\frac{\|\mathbf{I}_{xy}-\mathbf{p}_z\|^2}{2\tau^2})$, the feature recognition can be enhanced by adjusting τ according to the selected images. When the weighted image pixel $\mathbf{I}^z(x,y)<=\delta$, the value is set as 0, where δ is freely defined according to the object. Finally, the texture feature of each feature point is calculated by

$$\mathbf{TF}(\mathbf{p}_z)=\mathbf{DH}_{R,L}(\mathbf{BP}_{R,L}^{\mathbf{I}^z}) \tag{4}$$

where $\mathbf{DH}(\cdot)$ represents the statistics of the distribution histogram. Accordingly, the extracted texture feature from the reference and sensed images are $\mathbf{TF}_{\mathbf{t}_m}$ and $\mathbf{TF}_{\mathbf{s}_n}$, respectively. The difference in texture feature $\mathbf{TF}_{mn}$ between the source and target points can be evaluated by

$$\mathbf{TF}_{mn}=\|\mathbf{TF}_{\mathbf{s}_n}-\mathbf{TF}_{\mathbf{t}_m}\|^2 \tag{5}$$

3.3. Building the matching mapping for the feature candidates

After extracting multiple features, our goal shifts to find corresponding feature candidates among different sources images, and then establish their transformation $\mathcal{T}$. We decompose the main task of the feature-based registration method into two sub-tasks: correspondence estimation and transformation updating. These two sub-tasks are executed alternatively and iteratively until it reaches a steady state. To this end, the Expectation–Maximization (EM) algorithm is employed to solve $\mathcal{T}$, which includes two alternating steps:

- Expectation step (**E**-step): guessing the values of parameters ("old" parameter values estimated in the previous iteration) used to compute posterior probability distributions of mixture components based Bayes rule (computation of **CE**);
- Maximization step (**M**-step): computing the "new" parameter values via minimizing the expectation of the complete negative log-likelihood function.

These two steps of EM algorithm correspond to the above two sub-tasks.

Sub-task 1: correspondence estimation

We consider the deformation recovery as a problem of clustering the kernel center to the target point set of arbitrary shapes. To this end, we first estimate the probability density distribution on target point set.

$$PDF(x)=\frac{1}{M}\sum_{m=1}^{M}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{\|x-\mathbf{t}_m\|^2}{2\sigma^2}\right] \tag{6}$$

where σ is the bandwidth of Gaussian kernel. Analogous to the classic mean shift algorithm [32], we consider the source point $\mathbf{s}_n$ as the center of kernel and iteratively searches for the matched target points in the target point set $\{\mathbf{t}_m\}_{m=1}^M$. Afterwards, the gradient of each source point as:

$$Grad(\mathbf{s}_n)=\frac{\sum_{m=1}^{M}\Phi_{mn}}{M\sigma^2}\left[\frac{\sum_{m=1}^{M}\Phi_{mn}\mathbf{t}_m}{\sum_{m=1}^{M}\Phi_{mn}}-\mathbf{s}_n\right] \tag{7}$$

where $\Phi_{mn}=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{\|\mathbf{s}_n-\mathbf{t}_m\|^2}{2\sigma^2}\right]$ and the value in bracket denotes the normalized probability density gradient (aka. mean shift vector). The shift direction always points to the direction in which the probability density increases fastest, and its step size is proportional to the gradient probability density. Essentially, it is a gradient ascent algorithm with adaptive step. In mean shift algorithm, this vector is set as the displacement vector of point $\mathbf{s}_n$ in the process of searching corresponding target points. The item $\frac{\sum_{m=1}^{M}\Phi_{mn}\mathbf{t}_m}{\sum_{m=1}^{M}\Phi_{mn}}$ is the centroid (i.e., center of mass) of kernel determined by σ. Whereas, in the application of retinal image registration, there are differences in the appearance of retinal images from different image modalities, the extracted feature points inevitably contain a certain proportion of outliers. Let its proportion in the target point set be ω, and it conforms to uniform distribution $U(0,M)$. Thus the centroid of inliers determined by ω and σ as

$$\hat{\mathbf{s}}_n=\frac{(1-\omega)\sum_{m=1}^{M}\Phi_{mn}\mathbf{t}_m}{(1-\omega)\sum_{m=1}^{M}\Phi_{mn}+\frac{\omega}{M}} \tag{8}$$

Herein, each feature point also has texture feature and shape context feature. Thereby, the product of two radially symmetric Gaussian kernels is employed to define the multivariate kernels [32]

$$\Theta_{mn}=\frac{1}{2\pi\sigma\beta}exp\left[-\left(\frac{\mathbf{SC}_{mn}}{2\sigma^2}+\frac{\mathbf{TF}_{mn}}{2\beta^2}\right)\right] \tag{9}$$

where σ^2 and β^2 are affect bandwidths of the two kernels in their feature spaces, respectively. $\mathbf{SC}_{mn}+\mathbf{TF}_{mn}$ is called multiple features, which effectively merges texture and geometric structure information to enhance the recognition of feature points and the reliability of correspondence estimation under the interference of a large number of outliers. Therefore, the single kernel in Eq. (8) is replaced by multivariate kernels Θ.

$$\hat{\mathbf{s}}_n=\frac{(1-\omega)\sum_{m=1}^{M}\Theta_{mn}\mathbf{t}_m}{(1-\omega)\sum_{m=1}^{M}\Theta_{mn}+\frac{\omega}{M}} \tag{10}$$

The posterior probability matrix **CE** (**E**-step) of point $\hat{\mathbf{s}}_n$ in each iteration is estimated via Bayes law as

$$\mathbf{CE}_{mn} = \frac{(1-\omega)\sum_{m=1}^{M}\Theta_{mn}}{(1-\omega)\sum_{m=1}^{M}\Theta_{mn} + \frac{\omega}{M}} \tag{11}$$

Sub-task 2: Transformation updating

Assuming that the source point set $\{\mathbf{s}_n\}_{n=1}^{N}$ can be mapped point-wisely to the target point set $\{\mathbf{t}_m\}_{m=1}^{M}$ via spatial deformation $\mathcal{T}$. The transformation mapping for the alignment is dependent by an implicit variable $\vartheta = \{\mathcal{T}, \sigma^2\}$. For non-rigid deformation, the mapping $\mathcal{T}$ has infinite possibilities, and thus resulting the displacement field recovery be ill-posed. By the motion coherent theorem, the mapping $\mathcal{T}$ is smoothing such that for the closing points, they are moving to have the same displacement. The spatial transformation $\mathcal{T}$ is penalized by l_2-norm, $G(\mathcal{T}) = \|\mathcal{T}\|_2^2$, to preserve global motion consistency.

Additionally, it is hoped that these controlled inliers $\hat{\mathbf{s}}_n$ can thus be close to their corresponding points $\mathcal{T}(\mathbf{s}_n)$, and correspondingly drag the outliers around them, resulting in the latter drifts are scattered in a reasonable position. To this end, the dual geometric constraints are built to overcome over regularization caused by a strong global deformation G, as well as preventing over-fitting caused by local over-constraint,

$$\Re(\mathcal{T}) = \frac{\lambda}{2}\sum_{n=1}^{N}\|\hat{\mathbf{s}}_n - \mathcal{T}(\mathbf{s}_n)\|_2^2 + \frac{\eta}{2}G(\mathcal{T}) \tag{12}$$

According to Riesz representation theorem, non-rigid space transformation can be defined in reproducing kernel Hilbert space uniquely determined by Gaussian radial basis function. The kernel function is defined by using the Gram matrix composed of the spatial coordinates of the matrix. $\Gamma_{n_1 n_2} = \exp(-\frac{1}{2\phi^2}\|\mathbf{s}_{n_1} - \mathbf{s}_{n_2}\|^2)$, where $n_1, n_2 \in [1, N]$, and the constant ϕ controls the degree of spatial smoothness. Therefore, the non-rigid transformation function is formed as

$$\mathcal{T}(\mathbf{S}) = \mathbf{S} + \Gamma W \tag{13}$$

where Γ is a $M \times M$-dimensional positive definite matrix, W is the $N \times 2$-dimensional deformation coefficient matrix.

A reliable displacement direction yields a large expectation of probabilities on account of the probability density function. After that, the solution of the transformation updating is obtained by maximizing a likelihood function, or equivalent to minimizing the negative log-likelihood function. It is formulated as:

$$Q(\vartheta) = -\sum_{m=1}^{M} ln\left[(1-\omega)\sum_{n=1}^{N}\frac{1}{N}\Theta_{mn} + \frac{\omega}{M}\right] - \Re(\mathcal{T}) \tag{14}$$

Taking the upper bound of the energy function (14) and ignoring the irrelevant terms of ϑ, the posterior expectation of the complete-data log-likelihood is obtained via Jensen's inequality. Therefore, the maximization step (**M**-step) is achieved by minimizing the logarithmic posterior of the complete-data log-likelihood. The energy function is rewritten as

$$\begin{aligned} Q(W, \sigma^2, \omega) = {} & \frac{1}{2\sigma^2}\sum_{m=1}^{M}\sum_{n=1}^{N}\mathbf{CE}_{mn}\left\|\mathbf{T} - (\mathbf{S} + \Gamma W)_n\right\|^2 + \frac{\lambda}{2}tr(\mathcal{S}\mathcal{S}') \\ & + \frac{\eta}{2}tr(W'\Gamma W) + \mathbf{U}log\sigma^2 + \mathbf{U}log(1-\omega) + (M \\ & - \mathbf{U})log\omega \end{aligned} \tag{15}$$

where $\mathcal{S} = \mathbf{CE}\cdot\mathbf{T} - (\mathbf{S} + \Gamma W)$, $\mathbf{U} = \sum_{m=1}^{M}\sum_{n=1}^{N}\mathbf{CE}_{mn} \leqslant N$ (with $\mathbf{U} = N$ only if $\omega = 0$). The deformation coefficient can be obtained by partial derivation.

$$\begin{aligned} W = {} & (d(\mathbf{CE1})\Gamma + \eta\sigma^2\mathbf{I} + \lambda\sigma^2\Gamma)^{-1}(\mathbf{CE}\cdot\mathbf{T} - d(\mathbf{CE1})\mathbf{S} + \lambda\sigma^2(\mathbf{CE} \\ & \cdot\mathbf{T} - \mathbf{S})) \end{aligned} \tag{16}$$

where $d(\cdot)$ refers to matrix diagonalization. Similarly, σ^2 and ω are obtained by

$$\begin{aligned} \sigma^2 &= \frac{tr(\mathbf{T}'d(\mathbf{CE}'\mathbf{1})\mathbf{T}) - 2tr(\mathcal{T}(\mathbf{S})'\mathbf{CE}\cdot\mathbf{T})}{\mathbf{U}} + \frac{tr(\mathcal{T}(\mathbf{S})'d(\mathbf{CE1})\mathcal{T}(\mathbf{S}))}{\mathbf{U}} \\ \omega &= 1 - \frac{\mathbf{U}}{M} \end{aligned} \tag{17}$$

Subsequently, the obtained W, σ^2 and ω are substituted into the next iteration. The coordinates of transformed point set $\mathbf{S}^*$ are updated by $\mathcal{T}(\mathbf{S})$. Until the maximum number of iterations is reached or Eq. (15) converges, the point set registration is completed and the transformed feature source point set $\mathbf{S}^*$ is obtained.

3.4. Image registration

After obtaining a reliable correspondence $\Omega^* = \{\mathbf{S}, \mathbf{S}^*\}$, we are left to realize image registration via the correspondence. This study hopes that all inliers will not be disturbed by outliers and are precisely aligned with each other, while the outliers can be drifted to a reasonable position accordingly, exactly making the non-overlapping area of the grid image spread out well for guiding the resampling of the sensed image $\mathbf{I}^{\mathbf{S}}$. The backward approach is utilized to build image transformation based on the thin-plate splines (TPS) [21]. The detail of the image transformation is introduced in [4], which can obtain the transformed image $\mathbf{I}^{\mathbf{t}'}$.

Ultimately, the fusion of multi-source fundus images is a breeze, we can directly superimpose the reference and transformed images to obtain the fusion image.

3.5. Implementation details

3.5.1. Parameter setting

- the current number of iterations *iter* is initially set as 1 and incremented by 1 each time the iteration is executed, the maximum number of iterations $iter_{\max} \approx 60$;
- the initial threshold φ_0 and the reliable threshold φ^* are set to 1 and 2; the step parameter $\ell = (1.8 - 1.2)/(iter_{\max}/p)$, 1.8 and 1.2 are two moderate thresholds, where 1.8 is a good threshold with 95% inlier rate, 1.2 is a suitable threshold with 50% inlier rate; p is the period of the adjustable threshold update and set to 5;
- the values of *Ra* and *Ta* in shape context feature are set as 12 and 5 [33]; ρ_r and ρ_t are set as 1 and 1 [28];
- the radius R and neighbor points L of local binary pattern are set as 1 and 8 respectively;
- the bandwidth $\beta$is adjusted by the deterministic annealing scheme, i.e., $\beta = e^{\frac{1}{\mathcal{J}}}$, where the temperature parameter $\mathcal{J} = -\frac{iter}{5}$;
- the smoothing parameter ϕ is set to 2;
- the regularization coefficients λ and η are initially set to 2 and 3, updated by $\lambda = \mathcal{B}\lambda, \eta = \mathcal{B}\eta$, where $\mathcal{B} = \frac{(iter_{\max}^4 - iter + 1)^{1/144}}{iter_{\max}}$;
- the deformation parameter W and the outlier weighting parameter ω are initially set to 0 and 5, and then updated by Eqs. (16) and (17).

3.5.2. Algorithm pseudocode

The proposed algorithm is summarized by pseudocode, as shown in Algorithm 1.

Algorithm 1: Fully automatic multi-source retinal fundus image registration via feature-guided and dual structural preservation

input : a multi-source image pair $\mathbf{I^S}$ and $\mathbf{I^R}$

Construct the kernel matrix Γ;
Initialise φ_0 , φ^*, ω, W, λ, η, $iter$ and $iter_{\max}$;
Extract SIFT feature point sets $\mathbf{S}$ and $\mathbf{T}$ using φ_0 threshold;
repeat
  E-step:
  Calculate the feature difference matrices $\mathbf{SC}_{mn}$ and $\mathbf{TF}_{mn}$ by Eq. (3) and Eq. (5);
  Calculate $\mathcal{J}$ and $\mathcal{B}$ according to the current number of iterations;
  Compute the posterior probability matrix $\mathbf{CE}_{mn}$ by Eq. (11);
  Update the threshold $\varphi = \varphi + \ell$;
  M-step:
  Update W, σ and ω by Eq. (16) and Eq. (17);
  Update the location of $\mathbf{S}$ using Eq. (13);
  Update the annealing parameters by $\lambda{=}\mathcal{B}\lambda, \eta{=}\mathcal{B}\eta$;
until *reach* $iter_{max}$ *or Eq. (15) converges*;
Compute the transformed image $\mathbf{I^{t'}}$ using thin plate spline interpolation.

output : the transformed image $\mathbf{I^{t'}}$

4. Experiments and results

4.1. Dataset and experimental setting

The performance of the proposed method is conducted on three types of data: (i) 100 multi-source retinal fundus image pairs with minor displacement (named **Data i**); (ii) 100 multi-source retinal fundus image pairs with larger displacement or severe retinopathy lesions (**Data ii**); (iii) 20 image pairs involve a deliberate image impairment with additive different levels of Gaussian noises (**Data iii**). The resolutions of each image in the dataset are from 378×317 to 1703×1785 and these image pairs suffer rather different intensity profiles.

We evaluate the performance of the proposed method via comparing against seven feature-based methods including CPD [19], PR-GLS [22], GLMDTPS [20], MIF-RIRM [18], SIFT [15], SURF-PIIFD-RPM [26] and URSIFT-PIIFD-AGMM [27], which use their default parametric settings. The experiments are implemented in MATLAB 2019a on a Desktop PC with a 3.60-GHz Intel Core CPU and 16-GB RAM.

4.2. Evaluation criteria

To evaluate the registration results, a reliable and fair evaluation criterion is required to measure the performance of the aforementioned registration methods. Herein, we first manually select the ground truth, which includes 10–15 point pairs (i.e., landmarks) that are located in the obvious and easily identified places, such as the intersection of blood vessels, macular, optic papilla or the location of the lesion. Then, we calculate the Euclidean distance between the target feature points in the reference image and the corresponding feature points in the transformed image. Ultimately, we compute the root mean squared error (RMSE), mean absolute error (MAE), the median error (MEE) and the standard deviation of distance (STD) between the aforementioned two point sets. For their specific mathematical expressions, please refer to [33].

4.3. Ablation study

To verify the performance for each component of the proposed method, an ablation study is conducted by 20 multi-source retinal image pairs. Table 2 demonstrates the ablation investigation on the

Table 2
Ablation study of different components. '✓' denotes that the correspond component appears in the framework, '×' denotes that the fixed threshold, single feature description and single constraint are used in each step of registration process, respectively.

Case Index	1	2	3	4	5	6	7
Component i	✓	✓	×	✓	×	×	✓
Component ii	✓	×	✓	×	✓	×	✓
Component iii	×	✓	✓	×	×	✓	✓
RMSE	5.7673	8.6907	4.9708	14.7796	11.346	12.5256	**2.7430**
MAE	7.4693	10.7707	6.2924	18.2012	13.5176	16.8475	**3.4014**
MEE	2.2442	5.0024	1.9154	4.4355	2.5016	3.0191	**0.7044**
STD	4.7139	5.928	5.529	15.6712	16.1125	17.2939	**1.1317**

Table 3
Experimental statistics on the two types of the data involving **Data i** and **Data ii**. Average values of RMSE, MAE, MEE, STD and Run times for eight methods including (a) our method (b) CPD [19], (c) SIFT [15], (d) PR-GLS [22], (e) GLMDTPS [20], (f) MIF-RIRM [18], (g) SURF-PIIFD-RPM [26] and (h) URSIFT-PIIFD-AGMM [27]. Success rate [26] is recorded, which denotes the percentage of successful image pairs registration (inaccuracy: MAE $\leqslant$10 and MEE $>$ 1.5, acceptable: MAE $\leqslant$10 and MEE $\leqslant$1.5). The best performance is highlighted in bold.

Method		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Data i	RMSE	**2.09**	9.58	24.93	5.41	19.06	24.88	4.64	5.95
	MAE	**2.44**	12.26	32.39	6.83	23.50	31.29	5.86	7.34
	MEE	**0.98**	4.19	6.64	1.64	6.67	5.29	1.40	2.15
	STD	**1.48**	8.02	17.55	4.37	11.47	8.61	2.55	4.03
	Success rate	**96**%	68%	23%	82%	29%	75%	94%	91%
	Runtime (s)	3.72	0.63	**0.61**	11.45	1.05	1.64	7.98	7.62
Data ii	RMSE	**4.85**	12.07	76.91	15.82	28.15	29.72	4.93	7.24
	MAE	**5.91**	13.79	90.46	20.23	35.43	36.47	6.09	8.91
	MEE	**1.61**	4.34	25.38	5.30	8.08	9.14	1.77	3.10
	STD	4.51	8.55	14.08	9.13	12.88	14.96	**3.25**	6.36
	Success rate	91%	41%	7%	69%	17%	61%	**93**%	88%
	Runtime (s)	4.38	**0.69**	**0.69**	19.31	1.35	1.84	8.67	10.85

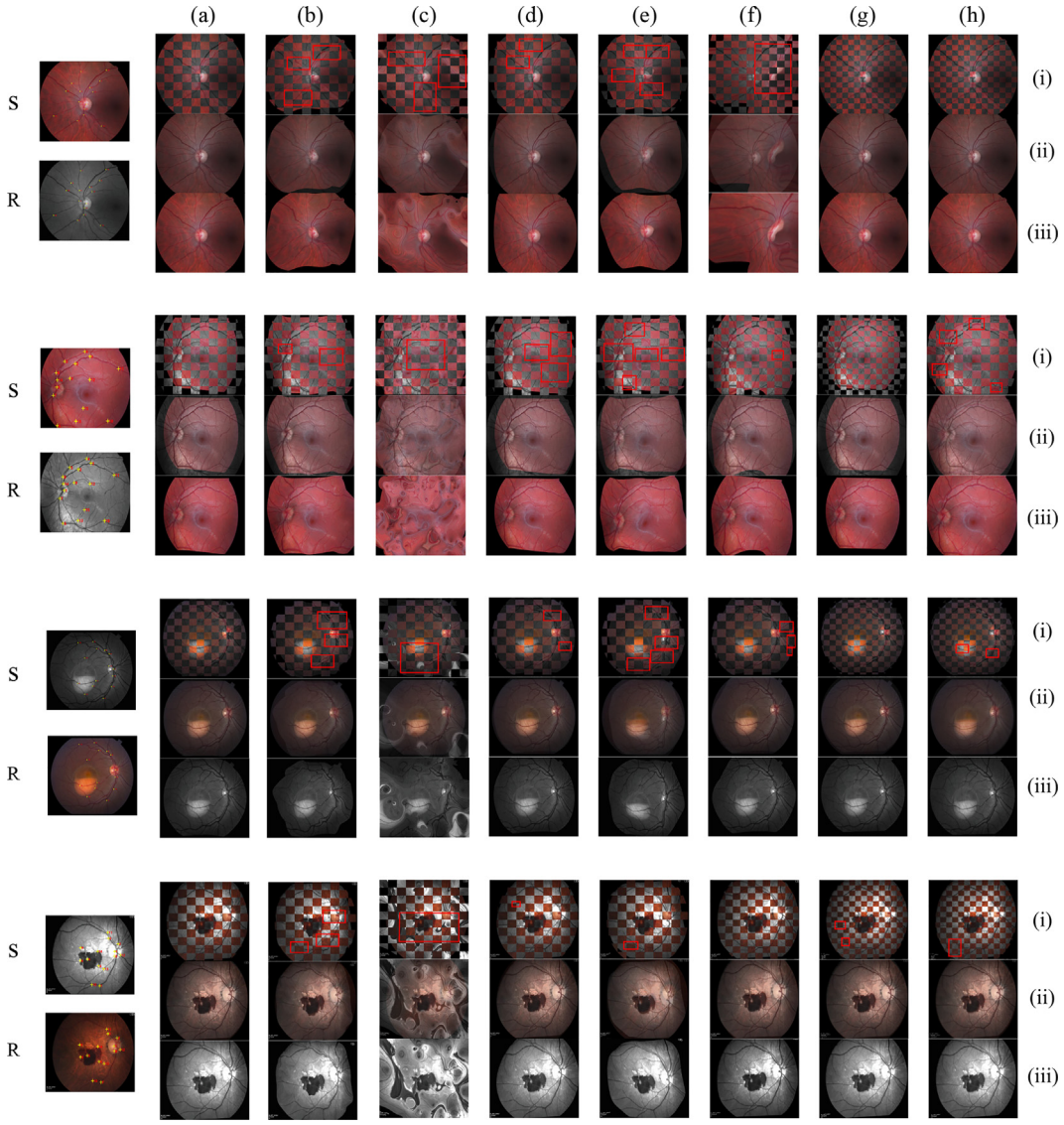

Fig. 5. Image registration results on four typical multi-source retinal fundus image pairs with a minor displacement among different sources. (a) our method (b) CPD [19], (c) SIFT [15], (d) PR-GLS [22], (e) GLMDTPS [20], (f) MIF-RIRM [18], (g) SURF-PIIFD-RPM [26] and (h) URSIFT-PIIFD-AGMM [27]. S and R are the sensed images and reference images, respectively. (i), (ii) and (iii) demonstrate the checkboard for alternately displaying the reference images and the transformed images, fusion images and transformed images, respectively. The registration errors are highlighted using the red rectangles in the checkboard images.

effects of the adjustable threshold selection strategy (component i), the multiple correspondences estimation model (component ii) and dual geometric structures constraints (component iii). First, to prevent false inlier matches, component (i) is adopted for maximizing the number of reliable inliers and making reasonable use of outliers as the number of iterations increases. Leveraging the relevant feature outliers from the entire image can contribute to the registration precision, since the extracted outlier pairs can be utilized as control points in building image transformation, and help to build a coarse to fine transformation. Second, the severe non-rigid deformation occurs during fundus imaging, single feature descriptors or constraints cannot guarantee a perfect transformed image, especially when one point is mismatched or several points need to move in different directions. Therefore, component (ii) and component (iii) are employed for improving the performance of image registration. The existence of each component is beneficial to improve the overall performance of registration, otherwise, using the fixed threshold, single feature and single constraint will degrade performance.

4.4. Results on multi-source retinal fundus images

The experimental results on **Data i** are reported in the first row Table 3. The intuitive results of four representative image pairs for CPD [19], SIFT [15], PR-GLS [22], GLMDTPS [20], MIF-RIRM [18], SURF-PIIFD-RPM [26], URSIFT-PIIFD-AGMM [27] and the proposed method are illustrated in Fig. 5. Experimental results demonstrate that the proposed method has excellent performance as well as the three important components, including (i) adjustable threshold selection strategy; (ii) multiple features; (iii) dual constraints, are reasonable. SIFT without components (i, ii and iii) performs not well and fails registration in most cases. It first uses a default threshold to extract insufficient feature points and then employs the strategy of nearest-neighbors distance ration to perform feature matching. However, the extraction operation may erroneously miss some inliers that are eliminated by the fast yet inaccurate strategy. CPD without components (ii and iii) only uses a single Euclidean distance feature to evaluate the one to many fuzzy correspondences, which will make the initial angle deviation between the source point set and the target point set at the beginning of registration. GLMDTPS and PR-GLS all although use multiple features to improve registration accuracy. GLMDTPS without component (iii) does not model the outliers present in the image and is sensitive to outliers, PR-GLS without component (iii) employs the rotation invariant shape context and inconsistent optimization processes to affect algorithm performance. MIF-RIRM without component (i) performs better due to the use of multiple features and dual constraints as well, the defect of the method is that the number of feature points extracted is fixed, resulting in more dubious estimation in areas with large differences in appearance. SURF-PIIFD-RPM and URSIFT-PIIFD-AGMM also perform promising, although our three components are not included in their frameworks, they use robust feature descriptors and conduct outliers rejection. The small flaw of these two methods is that they may not adequately utilize the image information due to a large number of outliers are removed.

The experimental results on **Data ii** are listed in the second row of Table 3. The intuitive results of four typical image pairs for PR-GLS [22], MIF-RIRM [18], SURF-PIIFD-RPM [26], URSIFT-PIIFD-AGMM [27] and our method are demonstrated in Fig. 6. The reason for choosing these four methods is that they performed better in the previous experiment. As can be seen from Fig. 6, the proposed method and SURF-PIIFD-RPM can generate quite a lot of correct matches, and the alignment results are almost perfect, even in the case of large-angle rotation or severe retinopathy. This can be seen from the seams of vessels in the checkerboard images.

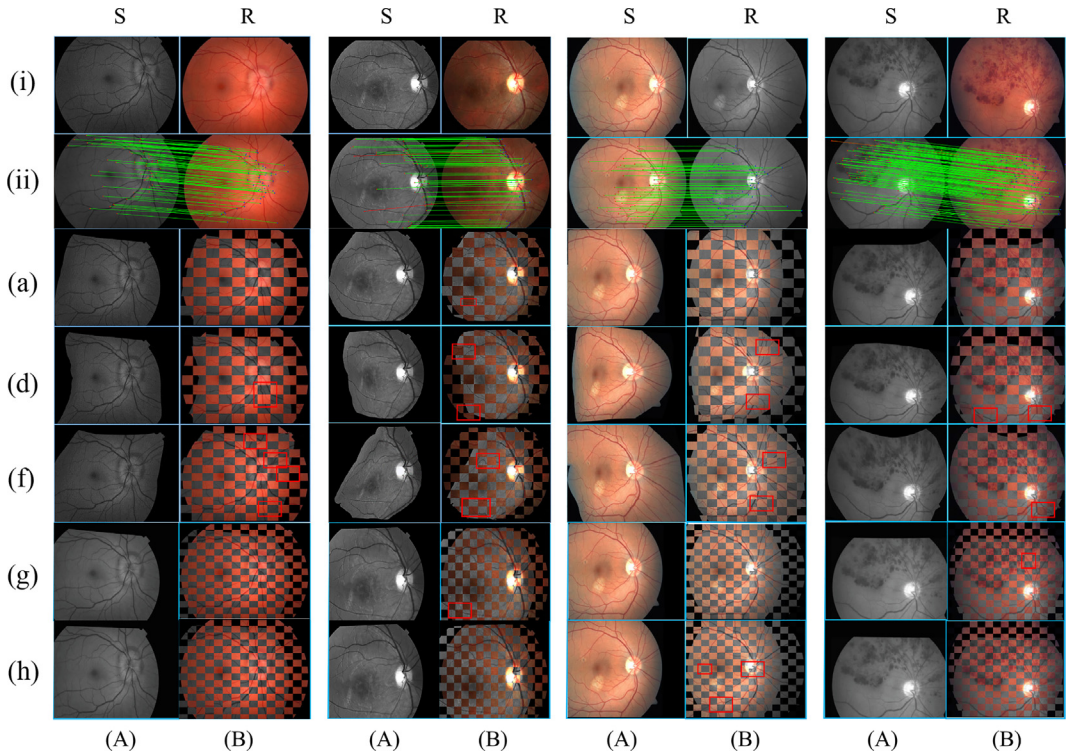

Fig. 6. Experiments on multi-source retinal fundus image registration with large displacement or severe retinopathy. (a) our method, (d) PR-GLS [22], (f) MIF-RIRM [18], (g) SURF-PIIFD-RPM [26] and (h) URSIFT-PIIFD-AGMM [27]. R and S denote the reference and sensed images in row (i), respectively. The (ii) row displays the feature matching results, where the correctly preserved matches are denoted by green lines; the false preserved matches are denoted by red lines. The (A) and (B) columns show the transformed images. and checkboard.

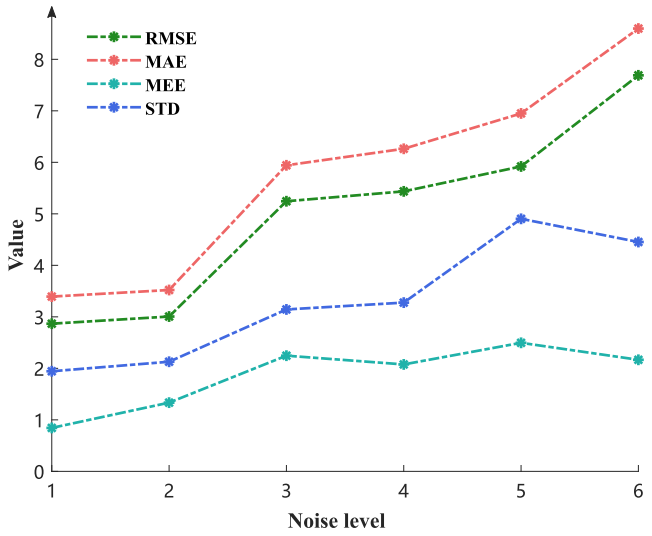

Fig. 7. Numerical results for **Data iii** with additive different levels (Level 1 to Level 6) of Gaussian noises.

Whereas the methods of PR-GLS and MIF-RIRM completely fail on most image pairs. URSIFT-PIIFD-AGMM seeks out relatively few corresponding points since the large-angle rotation or severe retinopathy, the performance is slightly worse than SURF-PIIFD-RPM. These results also justify the reasonability of incorporating multiple features, dual geometric structure constraints and the adjustable threshold selection strategy in the formulation.

We added different scales of Gaussian white noises to each fundus image for testing the robustness of the proposed method on **Data iii**. Specifically, we conducted deliberate image impairment and synthetically added noises with six levels (Level 1 to Level 6). Level 1 to Level 6 denote that the variances of Gaussian noise range from 0.01 to 0.1 in double steps. The numerical results of this experiment are depicted in Fig. 7, and two visualization examples are shown in Fig. 8. The experiment results indicate that our method consistently outperforms the comparison methods that are without adding Gaussian noises in most circumstances. It implies that our method is insensitive to noise and can tolerate a certain range of deliberate image impairment.

5. Discussion and conclusion

In this paper, a method for multi-source retinal fundus image registration via feature-guided and dual structural preservation is proposed. The method is demonstrated to achieve superior registration performance. The main merits of the method are summarized as follow: (i) an adjustable threshold selection strategy; (ii) a multiple correspondences estimation model and (iii) mean shift and the Tikhonov regularization based dual geometric structure constraints. In summary, the first merits provide sufficient and reliable inliers for the successive steps, the correspondence estimation provides a reliable correspondence matrix for the transformation updating, and the dual constraints pretty maintain the point structure in the process of alignment for ensuring the point moves in the right direction to achieve perfect alignment. The ablation study further proves that each component plays an important role in the whole image registration process, as evidenced by the significant difference after turning off either component. Compared to the performances of the proposed method with the seven registration methods, the proposed method shows the considerable performances.

However, our method exits some puny flaws, for instance, the selected SIFT algorithm may not be an optimal feature exactor, the selected features are not perhaps the most suitable feature combinations. We can provide mentality and direction to meet

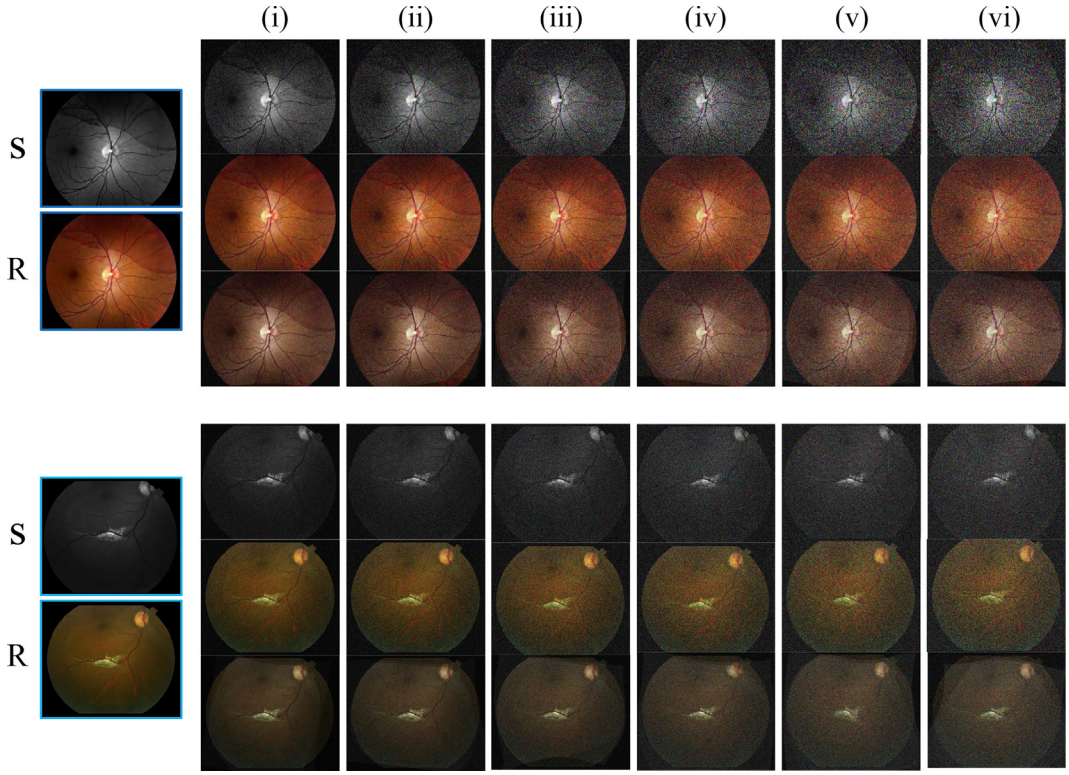

Fig. 8. Two examples for multi-source retinal fundus image registration with additive different levels (Level 1 (i) to Level 6 (vi)) of Gaussian noises. Each group is composed of three rows, the top two rows are the images after adding Gaussian noises, the bottom one is the fusion image.

some problems in the field. Furthermore, the proposed method is designed for handling the retinal image registration problem, which has the property of a few true feature correspondences and high percentages of false correspondences. Therefore, the proposed method can be applied to the monitoring of retinal diseases. Specifically, when the acquired retinal images are taken at different times or by different imaging modalities, the location of the lesions can be marked after registration to assist doctors in treatment to a large extend. In the future, we attempt to achieve the three-dimensional reconstruction of retinal images via the fusion images after accurate alignment of the image pairs for alleviating the workload of clinicians.

CRediT authorship contribution statement

Tingting Dan: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. **Yu Hu:** Writing - original draft. **Chu Han:** Writing - review & editing. **Zhihao Fan:** Visualization, Investigation. **Zhuobin Huang:** Visualization, Validation. **Bin Zhang:** Software. **Guihua Tao:** Software. **Baoyi Liu:** Resources. **Honghua Yu:** Resources. **Hongmin Cai:** Supervision, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank D. G. Lowe, J. Y. Ma, Y. Yang, A. Myronenko, S. Belongie and G. Wang for providing their implementation source codes. This greatly facilitated the comparison experiments. This work was partially supported by the National Natural Science Foundation of China (Grant No. 61771007, 61472145, 81870663), the Key-Area Research and Development of Guangdong Province (Grant No. 2020B010166002, 2020B1111190001), Guangdong Natural Science Foundation (Grant No. 2017A030312008), the Health & Medical Collaborative Innovation Project of Guangzhou City (Grant No. 201803010021, 202002020049), the Science and Technology Program of Guangzhou (Grant No. 202002030074).

References

[1] Y. He, W. Jiao, Y. Shi, B. Zhao, W. Zou, J. Lian, Y. Zhu, Y. Zheng, Segmenting diabetic retinopathy lesions in multispectral images using low-dimensional spatial-spectral matrix representation, IEEE Journal of Biomedical and Health Informatics 24 (2) (2020) 493–502, https://doi.org/10.1109/JBHI.2019.2912668.

[2] R. Bernardes, P. Guimarães, P. Rodrigues, P. Serranho, Fully-automatic multimodal co-registration of retinal fundus images, in: The International Conference on Health Informatics (ICHI), 2014, pp. 248–251, https://doi.org/10.1007/978-3-319-03005-0_63.

[3] L. Liang, W. Zhao, X. Hao, Y. Yang, K. Yang, L. Liang, Q. Yang, Image registration using two-layer cascade reciprocal pipeline and context-aware dissimilarity measure, Neurocomputing 371 (2020) 1–14, https://doi.org/10.1016/j.neucom.2019.06.101.

[4] Z. Yang, Y. Yang, K. Yang, Z. Wei, Non-rigid image registration with dynamic gaussian component density and space curvature preservation, IEEE Transactions on Image Processing 28 (5) (2019) 2584–2598, https://doi.org/10.1109/TIP.2018.2887204.

[5] F. Song, M. Li, Y. Yang, K. Yang, X. Gao, T. Dan, Small uav based multi-viewpoint image registration for monitoring cultivated land changes in mountainous terrain, International Journal of Remote Sensing 39 (21) (2018) 7201–7224, https://doi.org/10.1080/01431161.2018.1516051.

[6] A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration: A survey, IEEE Transactions on Medical Imaging 32 (7) (2013) 1153–1190, https://doi.org/10.1109/TMI.2013.2265603.

[7] Z. Yang, T. Dan, Y. Yang, Multi-temporal remote sensing image registration using deep convolutional features, IEEE Access 6 (2018) 38544–38555, https://doi.org/10.1109/ACCESS.2018.2853100.

[8] I. Rocco, R. Arandjelovic, J. Sivic, Convolutional neural network architecture for geometric matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (11) (2019) 2553–2567, https://doi.org/10.1109/TPAMI.2018.2865351.

[9] J. Ma, X. Jiang, A. Fan, J. Jiang, J. Yan, Image matching from handcrafted to deep features: A survey, International Journal of Computer Vision 62 (2020) 1–57, https://doi.org/10.1007/s11263-020-01359-2.

[10] J. Yang, X. Dong, Y. Hu, Q. Peng, G. Tao, Y. Ou, H. Cai, X. Yang, Fully automatic arteriovenous segmentation in retinal images via topology-aware generative adversarial networks, Neurocomputing 404 (2020) 14–25, https://doi.org/10.1016/j.neucom.2020.04.122.

[11] D. Mahapatra, B. Antony, S. Sedai, R. Garnavi, Deformable medical image registration using generative adversarial networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI), 2018, pp. 1449–1453, https://doi.org/10.1109/ISBI.2018.8363845.

[12] B. Zou, Z. He, R. Zhao, C. Zhu, W. Liao, S. Li, Non-rigid retinal image registration using an unsupervised structure-driven regression network, Neurocomputing 404 (2020) 14–25, https://doi.org/10.1016/j.neucom.2020.04.122.

[13] J. Lee, P. Liu, J. Cheng, H. Fu, A deep step pattern representation for multimodal retinal image registration, IEEE/CVF International Conference on Computer Vision (ICCV) 2019 (2019) 5076–5085, https://doi.org/10.1109/ICCV.2019.00518.

[14] Y. Wang, J. Zhang, M. Cavichini, D.U.G. Bartsch, W.R. Freeman, T.Q. Nguyen, C. An, Robust content-adaptive global registration for multimodal retinal images using weakly supervised deep-learning framework, IEEE Transactions on Image Processing 30 (2021) 3167–3178, https://doi.org/10.1109/TIP.2021.3058570.

[15] D.G. Lowe, Object recognition from local scale-invariant features, in: Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV), 1999, pp. 1150–1157, https://doi.org/10.1109/ICCV.1999.790410.

[16] C. Aguilera, F. Barrera, F. Lumbreras, A.D. Sappa, R. Toledo, Multispectral image feature points, Sensors 12 (12) (2012) 12661–12672, https://doi.org/10.3390/s120912661.

[17] H. Bay, A. Ess, T. Tuytelaars, L.V. Gool, Speeded-up robust features (surf), Computer Vision and Image Understanding 110 (3) (2008) 346–359, https://doi.org/10.1016/j.cviu.2007.09.014.

[18] D. Bi, R. Yu, M. Li, Y. Yang, K. Yang, S.H. Ong, Multiple image features-based retinal image registration using global and local geometric structure constraints, IEEE Access 7 (2019) 133017–133029, https://doi.org/10.1109/ACCESS.2019.2941256.

[19] A. Myronenko, X. Song, Point set registration: coherent point drift, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (12) (2010) 2262–2275, https://doi.org/10.1109/TPAMI.2010.46.

[20] Y. Yang, S.H. Ong, K.W.C. Foong, A robust global and local mixture distance based non-rigid point set registration, Pattern Recognition 48 (1) (2015) 156–173, https://doi.org/10.1016/j.patcog.2014.06.017.

[21] K. Yang, A. Pan, Y. Yang, S. Zhang, S.H. Ong, H. Tang, Remote sensing image registration using multiple image features, Remote Sensing 9 (6) (2017) 581–601, https://doi.org/10.3390/rs9060581.

[22] J. Ma, J. Zhao, A.L. Yuille, Non-rigid point set registration by preserving global and local structures, IEEE Transactions on Image Processing 25 (1) (2016) 53–64, https://doi.org/10.1109/TIP.2015.2467217.

[23] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (4) (2002) 509–522, https://doi.org/10.1109/34.993558.

[24] J. Ma, J. Jiang, J. Chen, C. Liu, C. Li, Multimodal retinal image registration using edge map and feature guided gaussian mixture model, Visual Communications and Image Processing (VCIP) 2016 (2016) 1–4, https://doi.org/10.1109/VCIP.2016.7805491.

[25] C. Liu, J. Ma, Y. Ma, J. Huang, Retinal image registration via feature-guided gaussian mixture model, Journal of the Optical Society of America A 33 (7) (2016) 1267–1266. doi:10.1364/JOSAA.33.001267..

[26] G. Wang, Z. Wang, Y. Chen, W. Zhao, Robust point matching method for multimodal retinal image registration, Biomedical Signal Processing and Control 19 (2015) 68–76, https://doi.org/10.1016/j.bspc.2015.03.004.

[27] H. Zhang, X. Liu, G. Wang, Y. Chen, W. Zhao, An automated point set registration framework for multimodal retinal image, in: 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 2857–2862, https://doi.org/10.1109/ICPR.2018.8545281.

[28] S. Zhang, Y. Yang, K. Yang, Y. Luo, S.H. Ong, Point set registration with global-local correspondence and transformation estimation, IEEE International Conference on Computer Vision (ICCV) 2017 (2017) 2688–2696, https://doi.org/10.1109/ICCV.2017.291.

[29] B. Remeseiro, A.M. Mendonca, A. Campilho, Objective quality assessment of retinal images based on texture features, International Joint Conference on Neural Networks (IJCNN) 2017 (2017) 4520–4527, https://doi.org/10.1109/IJCNN.2017.7966429.

[30] M. Omar, F. Khelifi, M.A. Tahir, Detection and classification of retinal fundus images exudates using region based multiscale lbp texture approach, in: 2016 International Conference on Control, Decision and Information Technologies (CoDIT), 2016, pp. 227–232, https://doi.org/10.1109/CoDIT.2016.7593565.

[31] R. Mehta, K. Egiazarian, Dominant rotated local binary patterns (drlbp) for texture classification, Pattern Recognition Letters 71 (99) (2016) 16–22, https://doi.org/10.1016/j.patrec.2015.11.019.
[32] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5) (2002) 603–619, https://doi.org/10.1109/34.1000236.
[33] T. Dan, Y. Yang, L. Xing, K. Yang, Y. Zhang, S.H. Ong, F. Song, X. Gao, Multifeature energy optimization framework and parameter adjustment-based nonrigid point set registration, Journal of Applied Remote Sensing 12 (3) (2018) 12–27, https://doi.org/10.1117/1.JRS.12.035006.

Tingting Dan is currently working toward the Ph.D. degree in the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. Her current research interests cover medical image processing, fMRI analysis and manifold learning.

Yu Hu is currently pursuing the Ph.D. degree with School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include clustering and medical image analysis.

Chu Han is now a postdoctoral fellow at the Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, under the supervision of Prof. Zaiyi Liu and Prof. Changhong Liang. He received his Ph.D. degree from the Chinese University of Hong Kong, under the supervision of Prof. Tien-Tsin Wong. His current research interests include medical image analysis, computer graphics, image processing, computer vision and deep learning.

Zhihao Fan is currently working toward the M.S. degree in the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His current research interests include medical image processing and super-resolution reconstruction.

Zhuobin Huang is currently working toward the M.S. degree in the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His current research interests include medical image processing and fMRI analysis.

Bin Zhang received the bachelor and master degree in School of Information Science and Engineering, from Shandong Normal University, Jinan, Shandong, in 2011 and 2015. He is currently working toward the Ph.D. degree in School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include multimedia, machine learning, medical image processing and bioinformatics.

Guihua Tao is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His current research interests cover deep learning and medical image processing.

Baoyi Liu is now a M.D. at the Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences; The Second School of Clinical Medicine, Southern Medical University, under the supervision of Prof. Honghua Yu. Her current research interests include retinal images and retinal diseases.

Honghua Yu is now a chief physician of Ophthalmology at the Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences. He completed postdoctoral fellow at the Schepens Eye Research Institute of Massachusetts Eye and Ear, Harvard Medical School, under the supervision of Prof. Dongfeng Chen. He received Ph.D. degree from the State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, under the supervision of Prof. Shibo Tang. His current research interests include retinal images and retinal diseases. Corresponding author of this paper.

Hongmin Cai is a Professor at the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. He received the B.S. and M.S. degrees in mathematics from the Harbin Institute of Technology, Harbin, China, in 2001 and 2003, respectively, and the Ph.D. degree in applied mathematics from Hong Kong University in 2007. From 2005 to 2006, he was a Research Assistant with the Center of Bioinformatics, Harvard University, and Section for Biomedical Image Analysis, University of Pennsylvania. His areas of research interests include biomedical image processing and omics data integration. Corresponding author of this paper.