

Received February 16, 2020, accepted February 19, 2020, date of publication February 27, 2020, date of current version March 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2976767

Non-Rigid Registration for Infrared and Visible Images via Gaussian Weighted Shape Context and Enhanced Affine Transformation

CHAOBO MIN¹, YAN GU², FENG YANG², YINGJIE LI³, AND WENJUN LIAN³

¹College of Internet of Things Engineering, Hohai University, Changzhou 213000, China

²North Night Vision Technology Company Ltd., Nanjing 211106, China

³North Information Control Research Academy Group Company, Ltd., Nanjing 211153, China

Corresponding author: Chaobo Min (chaobomin@outlook.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61901157.

ABSTRACT Image registration is a prerequisite for image fusion from multiple modalities, such as infrared (IR) and visible (VIS) images. Although there have been many various methods of image registration, non-rigid registration for IR and VIS images is still challenging due to large differences between IR and VIS images. In this work, a point feature-based method is proposed to improve the performance on non-rigid IR and VIS image registration. Firstly, a feature descriptor - Gaussian weighted shape context (GWSC) - is improved from shape context (SC) to fast extract matching point pairs from edge maps in IR and VIS images. With the set of matching point pairs, a Gaussian-field-based objective function is established to measure the distance between IR and VIS images. Then, the enhanced affine transformation (EAT) model is proposed to generalize affine model from linear to non-linear case and describe the regularity of global deformation between IR and VIS images. At last, the derivative of the distance measure is expressed with respect to the EAT model and thus, the optimal parameters are estimated by using the quasi-Newton method. The qualitative and quantitative comparisons demonstrate that the proposed method (GWSC-EAT) can be successfully applied to non-rigid registration of IR and VIS images and moreover, it is superior to the state-of-the-art methods on the accuracy and speed of non-rigid registration.

INDEX TERMS Image registration, non-rigid registration, shape context, non-linear transformation, infrared image.


I. INTRODUCTION

Image registration is an essential procedure in application for the fusion from different image sources, such as ultrasound and magnetic resonance image fusion [1] or infrared (IR) and visible (VIS) image fusion [2], which are very useful in medical imaging [3], machine vision [4], remote sense [5] and night vision [6], etc. The quality of image fusion depends highly on the accuracy of multimodal image registration [7]. The purpose of image registration is to determine the spatial transformation between images by using mutual feature of images. However, the extraction of mutual feature from multimodal images is a difficult task, because image feature can vary largely between multimodal images acquired with different acquisition parameters or systems. For example, the intensity-based feature of VIS images often loses in IR images. Therefore, multimodal image registration has always

been hot topic in the field of computer vision research. In this work, we mainly focus on IR and VIS image registration.

In many existing algorithms, multimodal image registration is considered as point set registration. The spatial transformation model aligning two images is estimated from feature point sets. Therefore, image registration is mainly divided into two parts: extraction of feature points and estimation of spatial transformation.

Although many registration methods do not need any explicit set of point correspondences, the sets of mutual feature points in multimodal images are still very essential for transformation estimation. Therefore, feature descriptors measuring the degree of point correspondence accurately are required. Intensity-based feature descriptors, such as corner feature [8], scale invariant feature transform (SIFT) [9], speeded up robust feature (SURF) [10] and histogram of gradient (HOG) [11], have been commonly applied for image registration. However, the different intensity distributions of IR and VIS images are challenging for intensity-based

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

feature descriptors. Structural feature descriptors are more applicable to image registration because the global structures of multimodal images to be aligned are similar [12]. Typically, shape context (SC) [13] is a widely applicable feature descriptor measuring the structure of point sets. Then, many SC-based feature descriptors, such as the inner distance SC (IDSC) [14], the coherent distance SC (CDSC) [15], the normalized weighted SC (NWSC) [16] and the rotation invariant SC (RISC) [17], have been improved from the original SC and successfully applied to shape matching or point set registration. But for IR and VIS image registration, they are not robust enough due to missing and deformed structures in IR and VIS images.

Estimation of transformation can be considered as an optimization procedure. Spatial transformation parameters are optimized to reduce the displacement between IR and VIS images to be aligned. Hence, a transformation model describing the pattern of deformation between IR and VIS images more accurately is more helpful for image registration. Linear transformation model, such as the affine model [18], cannot produce accurate alignment when there exists the anisotropy of deformation between images in numerous applications, especially in multi-sensor image fusion. Thus, many non-linear transformation models, such as the thin-plate spline (TPS) model [19], the B-splines model [20] and the model within reproducing kernel Hilbert space (RKHS) [21], have been developed for non-rigid registration. Something the above non-linear models have in common is that they all require control points. The transformation parameters are optimized in local neighborhoods of control points. Thus, the transformation models with control points are good at describing the pattern of local deformation nearby control points. It is obviously that control points have great impact on image registration. However, the optimal selection of control points is difficult to be determined, because the quantity and distribution of control points both affect the performance of estimation of spatial transformation. Therefore, the uncertainty of control points can result in inaccurate estimation of transformation parameters and increase computation complexity. In summary, we believe that the transformation models with control points rely heavily on local feature so that the performance of image registration is limited.

In this work, a structural feature descriptor - Gaussian weighted shape context (GWSC) - is developed to extract matching point pairs from edge maps in IR and VIS images. Non-rigid image registration is then transformed into point set registration. Meanwhile, the enhanced affine transformation (EAT) model consisting of affine model and polynomial model is proposed to describe the regularity of global deformation between IR and VIS images. In our method of non-rigid image registration (GWSC-EAT), a Gaussian-fields-based distance measure is established with the EAT model and then simplified by point correspondence with GWSC. Finally, the optimal EAT model is estimated from matching point pairs by a strategy with coarse-to-fine optimization.

Our contribution in this paper includes the following two aspects. Firstly, the feature descriptor GWSC is improved from SC in order to achieve accurate and fast matching point extraction from IR and VIS images. Secondly, the EAT model is proposed to reduce the dependence of non-rigid image transformation on local feature and improve the global accuracy of non-rigid image registration. Compared to previous approaches, it can increase the accuracy and speed of IR and VIS image registration. Hence, the proposed GWSC-EAT is able to improve the stability of IR and VIS image fusion systems.

In fact, our method is inspired by the non-rigid registration method proposed in [21] (namely RGF). Therefore, the regularized Gaussian field criterion is chosen as an objective function in our method. However, there are mainly two differences between the proposed method and RGF: 1) feature points was extracted by SC and the Hungarian algorithm [39] in RGF, while we improve the efficiency of feature point extraction via GWSC; 2) the RKHS-based transformation model used in RGF prefers to describe local deformation in neighborhoods of control points, while we develop the EAT model to describe global regularity of non-linear deformation, in order to improve the global accuracy of non-rigid image registration.

The rest of the paper is organized as follows. Section II reports background material and related works. Section III describes the algorithm of matching point extraction by GWSC. In Section IV, we present the EAT model and apply it to non-rigid registration of IR and VIS images. Section V shows the experiment results on real images and the analysis of the proposed method. Finally, Section VI presents the concluding remarks for our work.

II. RELATED WORKS

In many researches, multimodal image registration is formulated as an optimization problem, where an objective function is minimized with respect to a spatial transformation [12]. Objective function can be regarded as a distance measure, quantifying registration performance during optimization process. Transformation model is used to determine the search direction of optimization. Thus, two key points researchers mainly focus on are as follows: 1) find measures that can capture the similarity between multimodal images; 2) establish spatial transformation models describing the pattern of deformation between multimodal images. In this section, we mainly focus on the current researches of similarity measures since the popular transformation models were discussed in Section I.

Intensity-based measures have been used widely in image registration, which are defined directly on gray-level values of images. Mutual information (MI) [22] and normalized mutual information (NMI) [23] are typical intensity-based measures for image registration. They mainly rely on the assumption that the statistical regularity of intensity is similar between images to be aligned. MI and NMI have been successfully employed in various applications,

especially in mono-modal applications. However, because there are highly different intensity distributions in multimodal images, MI-based measures may not quantify the performance of alignment correctly [24]. Furthermore, because of non-convex objective functions, it is quite difficult to solve them quickly and accurately [25]. With the development of research, MI-based measures not only work directly with image intensity, but also with image gradient [26] and image patch [27].

In recent years, spectral methods have gained interest for image registration, which use spectral decomposition to study image data structures. They can be considered as the upgrading of intensity-based measures, providing a way for constructing high dimensional eigenspace by global intensity and structure of an image. In [28] and [29], multimodal images were represented by the first embedding coordinate of Laplacian eigenmaps and diffusion maps, respectively, where the L_1 or L_2 distance measures were used to describe the similarity between the structural representations of images. In [30] and [31], a joint graph of two images or shapes obtained by spectral decomposition was used for image matching and surface matching. In [12], graph Laplacian of an image was employed as a structure descriptor and the similarity between graph Laplacian eigenspaces of multimodal images was measured by Laplacian commutativity. Because the similarity measures constructed by spectral methods are on the basis of L_1 or L_2 distance measures, they are convex functions and helpful for optimization of image registration. However, the derivatives of such measures are computationally expensive since many parameters have to be optimized.

Feature-based measures have always been concerned by many researchers. Image features, which can reflect the correspondence between multimodal images, such as points, curves and surfaces [32], are extracted from images. The distance measures defined on the extracted features are then used to quantify the performance of image registration. Actually, point feature is more popular since curves and surfaces can be regarded as point sets. Hence, with feature-based measures, image registration is transformed into point set registration. A transformation model aligning two point sets is then applied to achieve image registration.

Feature-based measures can be formulated by L_2 loss criterion [33], L_2 -minimizing estimator (L_2E) [34], regularized Gaussian field criterion [21] and Gaussian mixture model (GMM) [35], [36], etc. These models ensure that feature-based measures are convex functions and have the advantage of computational convenience. Point correspondence is very essential to feature-based measures, because the distances between point pairs that do not match each other in fact can result in low accuracy of measurement. In some existing registration methods, such as iterated closet point (ICP) [37], robust point matching using TPS (TPS-RPM) [33], coherence point drift (CPD) [38], robust point matching using L_2E estimator (RPM- L_2E) [34] and image registration using RGF [21], the correspondence was iteratively solved by optimization with soft-assignment as initial matching.

However, because the feature-based measures with soft-assignment must be defined on all pairs of feature points, they are computationally expensive. Thus, some algorithms of point correspondence are usually employed to determine mutual point pairs from multimodal images before estimation of transformation models. For instance, in [17], point correspondence was implemented by the Hungarian method with RISC and then, the L_2 distance measure was defined on matching pairs of feature points so that its complexity was greatly reduced. Hence, we find that if point correspondence is accurate enough, it can simplify feature-based measures and ensure measure performance. In general, point correspondence consists of feature descriptors and bipartite graph matching. Such feature descriptors have been introduced in Section I. The usual algorithms of bipartite graph matching include the Hungarian algorithm [39], the deferred-acceptance algorithm [40] and the shortest augmenting path algorithm [41], etc.

Compared with intensity-based and spectral-based measures, feature-based measures and the corresponding derivatives are simpler in mathematical form. As a result, gradient-based numerical optimization technique is easily applied to the optimization for feature-based measures. This is very helpful to improve the accuracy and speed of estimation of transformation models.

III. MATCHING POINT EXTRACTION BY GWSC

In order to formulate image registration as an optimization problem, image registration can be considered as point set registration. In our work, point sets are extracted from edge maps of IR and VIS images. Hence, in this section, we present the Gaussian-fields-based distance measure and matching point extraction by GWSC.

A. GAUSSIAN-FIELDS-BASED DISTANCE MEASURE

The feature point sets extracted from the edge maps of IR and VIS images are represented as $R = \{\mathbf{r}_m\}_{m=1}^M$ and $S = \{\mathbf{s}_n\}_{n=1}^N$, $\mathbf{r}_m, \mathbf{s}_n \in \mathbb{R}^2$, respectively. In this work, the goal of registration is to align IR images to VIS images. VIS images are considered as reference images and IR images are considered as input images. $\varphi_{\mathbf{p}}$ is a transformation model with parameters \mathbf{p} , which maps points from IR image space to VIS image space. The solution of image registration is to estimate transformation parameters \mathbf{p} by aligning R to S with $\varphi_{\mathbf{p}}$. Estimation of transformation can be regarded as an optimization procedure. Therefore, an objective function, which is able to measure registration accuracy during registration quite precisely, is very essential.

We choose Gaussian fields to construct the objective function because it is continuously differentiable and could converge to global optimal solution quickly [21]. The objective function that can be optimized during registration is therefore

$$\min_{\mathbf{p}} E(\mathbf{p}) = \min_{\mathbf{p}} - \sum_{m=1}^M \sum_{n=1}^N C_{mn} \exp \left\{ -\frac{\|\mathbf{s}_n - \varphi_{\mathbf{p}}(\mathbf{r}_m)\|^2}{2\sigma_e^2} \right\} + \lambda S(\mathbf{p}), \quad (1)$$

where $\|\cdot\|$ denotes the L_2 norm, σ_e is a range parameter, \mathbf{C}_{mn} indicates the correspondence between points \mathbf{r}_m and \mathbf{s}_n . The first term describes the distance between two point sets, and the second term $S(\mathbf{p})$ provides some control over the transformation. $\lambda \in \mathbb{R}$ is a regularization penalty weight that balances the two terms.

\mathbf{C} can be considered as a binary correspondence matrix, where $\mathbf{C}_{mn} = 1$ if a point \mathbf{r}_m matches to a point \mathbf{s}_n , otherwise $\mathbf{C}_{mn} = 0$. Thus, \mathbf{C} shows the point correspondences between the point sets R and S . Apparently, the correspondence matrix \mathbf{C} determines the measuring accuracy of $E(\mathbf{p})$, because incorrect point correspondence may result in failure of the quantization of registration performance. Meanwhile, in a certain extent, the correspondence matrix \mathbf{C} determines the time complexity of optimization for the objective function (1). The iteration number of optimization is represented as a and the time complexity of optimization for (1) is $O(MNa)$ at least. Thus, the sizes of feature point sets R and S must be small enough to avoid high computational complexity for solving optimization problem. However, the global accuracy of image registration cannot be correctly quantified from too few feature points. Hence, if a set of matching point pairs $F = \{(\mathbf{r}_k, \mathbf{s}_k)\}_{k=1}^K$ can be precisely extracted from edge maps of IR and VIS images, the objective function can be greatly simplified and its measuring accuracy can be improved.

B. GAUSSIAN WEIGHTED SHAPE CONTEXT

SC is a feature descriptor which can describe the neighborhood structures of points well. Therefore point correspondence can be determined from edge maps by the similarity measure with SC. In this work, the edge maps of IR and VIS images are extracted by canny edge descriptor [42]. The similarity measure between two points is defined as

$$\mathbf{C}_{s_{ij}} = \frac{1}{2} \sum_{t=1}^T \frac{(S_t(\mathbf{b}_i^r) - S_t(\mathbf{b}_j^v))^2}{(S_t(\mathbf{b}_i^r) + S_t(\mathbf{b}_j^v))}, \quad (2)$$

where the edge point sets of IR and VIS images are represented as $\{\mathbf{b}_i^r\}_{i=1}^I$ and $\{\mathbf{b}_j^v\}_{j=1}^J$ ($\mathbf{b}_i^r, \mathbf{b}_j^v \in \mathbb{R}^2$) respectively. $S_t(\cdot)$ denotes the T -bin normalized histogram at an edge point,

$$S_t(\mathbf{b}_i) = \#\{\mathbf{q} \neq \mathbf{b}_i: (\mathbf{q} - \mathbf{b}_i) \in \text{bin}(t)\}, \quad (3)$$

where \mathbf{b}_i and \mathbf{q} both represent edge points of an image, $\mathbf{b}_i, \mathbf{q} \in \mathbb{R}^2$. The more details of calculating $\mathbf{C}_{s_{ij}}$ are shown in [13]. \mathbf{C}_s can be regarded as a cost matrix between the edge points of IR and VIS images. The lower the similarity measure between two edge points is, the more similar they are.

SC has good performance on point set registration, but it is often failed in IR and VIS image registration. Fig. 1 shows an example of extracting a pair of matching points from edge maps by SC. The point A is determined to be corresponding to the point B by SC, but actually the correspondence between the points A and C is true. The reason why there is such mismatch is that edge features of VIS images may be lost or deformed in IR images. In our work, two approaches

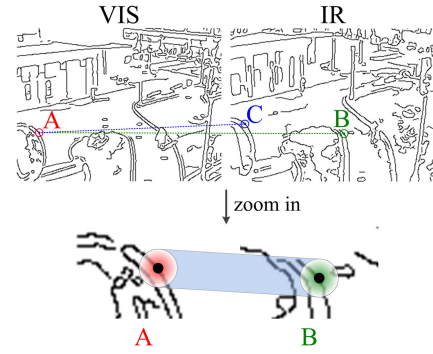


FIGURE 1. An example of point correspondence by the original SC.

are applied to improve in the robustness of SC for IR and VIS image registration. One is to measure the similarity between two points by using average SC in neighborhood, as shown in the bottom of Fig. 1, in order to increase the range of structural feature. The other one is to improve the distinction between non-corresponding points via the relative distance between edge points \mathbf{b}_i^r and \mathbf{b}_j^v . Therefore, a cost matrix with the feature descriptor GWSC can be written as follow,

$$\mathbf{C}_g = (\mathbf{D}_r^{-1} \mathbf{W}^r \mathbf{C}_s \mathbf{W}^v \mathbf{D}_v^{-1}) * \mathbf{W}^{rv}, \quad (4)$$

where \mathbf{W}^r , \mathbf{W}^v and \mathbf{W}^{rv} are the weight matrixes computed by using Gaussian kernel as follows: $\mathbf{W}_{il}^r = \exp\{-\varepsilon_r \|\mathbf{b}_i^r - \mathbf{b}_l^r\|^2\}$, $i, l \in \{1, 2, \dots, I\}$; $\mathbf{W}_{ju}^v = \exp\{-\varepsilon_v \|\mathbf{b}_j^v - \mathbf{b}_u^v\|^2\}$, $j, u \in \{1, 2, \dots, J\}$; $\mathbf{W}_{ij}^{rv} = \exp\{\varepsilon_{rv} \|\mathbf{b}_i^r - \mathbf{b}_j^v\|^2\}$. $\varepsilon_r, \varepsilon_v, \varepsilon_{rv} \in \mathbb{R}^+$ determine the range of interaction between points (i.e. neighborhood size). $\mathbf{D}_r = \text{diag}(\sum_{b=1}^I \mathbf{W}_{ib}^r)$ and $\mathbf{D}_v = \text{diag}(\sum_{b=1}^J \mathbf{W}_{jb}^v)$ are the diagonal degree matrixes of \mathbf{W}^r and \mathbf{W}^v , respectively. $*$ denotes Hadamard product. \mathbf{C}_g is an attribute matrix between edge points of IR and VIS images, where the lower $\mathbf{C}_{g_{ij}}$ denotes that the possibility of correspondence between two edge points \mathbf{b}_i^r and \mathbf{b}_j^v are greater.

C. EXTRACTION OF MATCHING POINT PAIRS

On the basis of cost matrix, the correspondence between two point sets can be determined by the algorithms of bipartite graph matching. However, it is not required for image registration to determine the maximum matching between all edge points in IR and VIS images. Furthermore, there must be incorrect point correspondence in the maximum matching due to missing boundaries between IR and VIS images. Therefore, the algorithms of bipartite graph matching, such as the Hungarian method, are not suitable for image registration. In addition, bipartite graph matching is computationally expensive since the numbers of edge points in real images are very large in most cases.

For achieving image registration, the major purpose of point correspondence is to extract matching point pairs from all edge points. Here, a fast algorithm is developed to extract matching point pairs with the cost matrix \mathbf{C}_g . It is outlined as follows.

Algorithm 1 Fast Extraction of Matching Point Pairs

Input: an pair of IR and VIS image, parameters $\varepsilon_r, \varepsilon_v$ and ε_{rv}

Output: the set of matching point pairs F

- 1 Use canny edge descriptor to extract edge point sets $\{\mathbf{b}_i^r\}_{i=1}^I$ and $\{\mathbf{b}_j^v\}_{j=1}^J$ from IR and VIS images;
- 2 Compute the GWSC cost matrix \mathbf{C}_g between $\{\mathbf{b}_i^r\}_{i=1}^I$ and $\{\mathbf{b}_j^v\}_{j=1}^J$ by (4);
- 3 Locate the minimum of each row in \mathbf{C}_g and a set of the corresponding column indices is indicated as M_v ;
- 4 Locate the minimum of each column in \mathbf{C}_g and a set of the corresponding row indices is indicated as M_r ;
- 5 According to M_v , a set of point correspondence can be written as $C_v = \{(\mathbf{b}_i^r, \mathbf{b}_q^v) | i \in \{1, 2, \dots, I\}, q \in M_v\}$;
- 6 Similarly, $C_r = \{(\mathbf{b}_p^r, \mathbf{b}_j^v) | j \in \{1, 2, \dots, J\}, p \in M_r\}$;
- 7 The set of matching point pairs can be determined by $F = \{(\mathbf{r}_k, \mathbf{s}_k)\}_{k=1}^K = C_v \cap C_r$.

Using Algorithm 1, the point pairs with the relatively lowest cost of GWSC are extracted as the feature points for the distance measure. Therefore, the algorithm of fast extraction can reduce incorrect point correspondence as far as possible. Moreover, it is obviously that the time complexity of the steps 3~7 is $O(IJ)$, which is far less than that of bipartite graph matching. In addition, F actually contains the point pairs that are most likely to match each other and thus, it indicates potential correspondence between IR and VIS images.

According to the set of matching point pairs F , the objective function (1) can be simplified by

$$\min_{\mathbf{p}} E_S(\mathbf{p}) = \min_{\mathbf{p}} - \sum_{k=1}^K \exp \left\{ - \frac{\|\mathbf{s}_k - \varphi_{\mathbf{p}}(\mathbf{r}_k)\|^2}{2\sigma_e^2} \right\} + \lambda S(\mathbf{p}), \quad (5)$$

where $(\mathbf{r}_k, \mathbf{s}_k) \in F$. Compared with (1), the simplified objective function IV) has lower computational complexity, because K is significantly less than $M \times N$ in general.

IV. ESTIMATION OF ENHANCED AFFINE TRANSFORMATION

Transformation model $\varphi_{\mathbf{p}}$ is a map function determining directly the accuracy of image registration. Meanwhile, $\varphi_{\mathbf{p}}$ indicates the pattern of deformation between two point sets (or two images). In this section, the EAT model is developed to describe the global regularity of non-rigid deformation between IR and VIS images. Then, a strategy with coarse-to-fine optimization is employed to estimate the parameters \mathbf{p} of the EAT model from the set of matching point pairs.

A. EAT MODEL

In general, IR and VIS images to be aligned are captured by image fusion systems with parallel optical axis of IR and VIS cameras. The deformation between IR and VIS images

is caused by the following factors: the displacement between IR and VIS cameras, the different lens of IR and VIS cameras and the different parameters of IR and VIS sensors, etc. Thus, the deformation between an image pair can be considered as a mixture of various regular patterns. On the basis of the above analyses, we believe that the global deformation between an IR and VIS image pair should exhibit a more complex regular pattern.

The affine transformation model is a typical model describing the global regularity of linear deformation between two images, where the spatial transformations of all points are conformed to a unified pattern. The affine model has the superiorities of easy implementation and low cost. Furthermore, a transformation model with global regularity has two advantages: 1) there is no requirement of control points; 2) the scale of spatial transformation is consistent for all points. Thus, it is helpful for non-rigid image registration to generalize the affine model from linear to non-linear case in order to accurately describe the complex regular pattern of global deformation between IR and VIS images without control points. This is also the motivation of the proposed EAT model.

$\mathbf{r}_k = [x, y]$ is a 1×2 dimensional coordinate vector and $\varphi_{\mathbf{p}}(\mathbf{r}_k) = [\hat{x}, \hat{y}]$ represents the mapping location of \mathbf{r}_k . The EAT model is formulated by

$$\varphi_{\mathbf{p}}(\mathbf{r}_k) = [x, y, 1] \mathbf{A}^T + [G(x, y, \boldsymbol{\alpha}), G(x, y, \boldsymbol{\beta})], \quad (6)$$

where the affine transformation matrix $\mathbf{A} = [s_x \cos \theta, -\sin \theta, t_x; \sin \theta, s_y \cos \theta, t_y]$, θ is angle of rotation, s_x and s_y are scaling coefficients, t_x and t_y are translation coefficients. $G(\cdot)$ is defined as follow:

$$\begin{cases} G(x, y, \boldsymbol{\alpha}) = \sum_{i=2}^5 \sum_{j=0}^i w_i \alpha_{i,j} x^j y^{i-j} \\ G(x, y, \boldsymbol{\beta}) = \sum_{i=2}^5 \sum_{j=0}^i w_i \beta_{i,j} x^j y^{i-j} \end{cases} \quad (7)$$

where $\alpha_{i,j}, \beta_{i,j} \in \mathbb{R}$, $w_i \in \mathbb{R}^+$. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are respectively the 18×1 dimensional parameter vectors of polynomial transformation. $\boldsymbol{\alpha} = [\alpha_{2,2}, \alpha_{2,0}, \alpha_{2,1}, \alpha_{3,3}, \alpha_{3,0}, \alpha_{3,2}, \alpha_{3,1}, \alpha_{4,4}, \alpha_{4,0}, \alpha_{4,3}, \alpha_{4,2}, \alpha_{4,1}, \alpha_{5,5}, \alpha_{5,0}, \alpha_{5,3}, \alpha_{5,2}, \alpha_{5,4}, \alpha_{5,1}]$. $\boldsymbol{\beta} = [\beta_{2,2}, \beta_{2,0}, \beta_{2,1}, \beta_{3,3}, \beta_{3,0}, \beta_{3,2}, \beta_{3,1}, \beta_{4,4}, \beta_{4,0}, \beta_{4,3}, \beta_{4,2}, \beta_{4,1}, \beta_{5,5}, \beta_{5,0}, \beta_{5,3}, \beta_{5,2}, \beta_{5,4}, \beta_{5,1}]$. The weight vector with w_i is defined as $\mathbf{w} = [w_2, w_2, w_2, w_3, w_3, w_3, w_3, w_4, w_4, w_4, w_4, w_5, w_5, w_5, w_5, w_5, w_5]$, which is used to balance the two terms of (6) and control $G(\cdot)$ in a reasonable range. Moreover, the weight vector is able to determine which terms in $G(\cdot)$ are calculated for the EAT model, in order to adjust the nonlinearity of the EAT model. The first term of (6) is the affine transformation model describing the pattern of linear deformation between two point sets, while the second term is a non-linear transformation model consisting of quadratic, cubic, quartic and quintic polynomial transformation models. Substituting (7) into (6), the EAT model becomes the

following matrix form:

$$\varphi_{\mathbf{p}}(\mathbf{r}_k) = \pi_k \mathbf{A}_e(\mathbf{p}), \quad (8)$$

where the 1×21 dimensional polynomial vector $\pi_k = [x, y, 1, x^2, y^2, xy, x^3, y^3, x^2y, xy^2, x^4, y^4, x^3y, x^2y^2, xy^3, x^5, y^5, x^3y^2, x^2y^3, x^4y, xy^4]$. $\mathbf{A}_e(\mathbf{p})$ is considered to be the EAT matrix with the parameter vector \mathbf{p} and defined as

$$\mathbf{A}_e(\mathbf{p}) = \left[\mathbf{A} \left[\alpha_w | \beta_w \right]^T \right]^T, \quad (9)$$

where the weighted parameter vectors of polynomial transformation are $\alpha_w = \mathbf{w}^T * \alpha^T$ and $\beta_w = \mathbf{w}^T * \beta^T$, respectively. The 41×1 dimensional parameter vector $\mathbf{p} = [\theta, s_x, s_y, t_x, t_y | \alpha | \beta]^T$. By using (8), the objective function IV) becomes:

$$\min_{\mathbf{p}} E_S(\mathbf{p}) = \min_{\mathbf{p}} - \sum_{k=1}^K \exp \left\{ - \frac{\|\mathbf{s}_k - \pi_k \mathbf{A}_e(\mathbf{p})\|^2}{2\sigma_e^2} \right\} + \lambda \text{tr} \left((\mathbf{p} - \mathbf{z})(\mathbf{p} - \mathbf{z})^T \right), \quad (10)$$

where the 41×1 dimensional vector $\mathbf{z} = [0, 1, 1, 0, \dots, 0]^T$, $\text{tr}(\cdot)$ denotes the trace. The second term of (10) describes the change extent of the parameter vector \mathbf{p} during optimization process. It is able to give the objective function penalty when the variation of \mathbf{p} is excessive.

B. OPTIMIZATION

It can be seen in (10) that the objective function with the EAT model is continuously differentiable with respect to the transformation parameter vector \mathbf{p} . Thus, the derivative of (10) is given by

$$\frac{\partial E_S(\mathbf{p})}{\partial \mathbf{p}} = \frac{1}{\sigma_e^2} \sum_{k=1}^K \frac{\partial \pi_k \mathbf{A}_e(\mathbf{p})}{\partial \mathbf{p}} (\pi_k \mathbf{A}_e(\mathbf{p}) - \mathbf{s}_k)^T \exp \left\{ - \frac{\|\mathbf{s}_k - \pi_k \mathbf{A}_e(\mathbf{p})\|^2}{2\sigma_e^2} \right\} + 2\lambda(\mathbf{p} - \mathbf{z}), \quad (11)$$

where the derivative of the EAT model can be written as follow:

$$\frac{\partial \pi_k \mathbf{A}_e(\mathbf{p})}{\partial \mathbf{p}} = \begin{bmatrix} -s_x x \sin \theta - y \cos \theta & x \cos \theta - s_y y \sin \theta \\ x \cos \theta & 0 \\ 0 & y \cos \theta \\ 1 & 0 \\ 0 & 1 \\ \mathbf{w}^T * \hat{\pi}_k^T & \mathbf{0} \\ \mathbf{0} & \mathbf{w}^T * \hat{\pi}_k^T \end{bmatrix}, \quad (12)$$

where $\hat{\pi}_k = [x^2, y^2, xy, x^3, y^3, x^2y, xy^2, x^4, y^4, x^3y, x^2y^2, xy^3, x^5, y^5, x^3y^2, x^2y^3, x^4y, xy^4]$, $\mathbf{0}$ represents 18×1 dimensional zero vector.

By using the derivative (11), gradient-based numerical optimization technique can be employed to determine the optimal transformation parameter vector \mathbf{p} . In this work, the quasi-Newton method is introduced to solve such optimization problem. However, the optimization procedure is

limited by local convergence, because of the following reasons: 1) the Gaussian-field-based objective function is convex only in the neighborhood of the optimal solution; 2) the convergence of the quasi-Newton method is susceptible to the choice of initial value. It is obviously that the chance of reaching the global optima is partly determined by the EAT model. The EAT model with low nonlinearity is able to prevent the objective function (10) from local convergence. But registration error may be large, because the EAT model with low nonlinearity is not accurate enough to describe the non-rigid deformation between two point sets. The EAT model with over high nonlinearity may not properly optimize the objective function, because it can increase the influence of initial error between IR and VIS images, resulting in deviation of gradient direction, as seen in (12). In the proposed EAT model, it is easy to adjust the orders of polynomial transformations by using the weight vector \mathbf{w} , in order to regulate the nonlinearity of the EAT model. Therefore, to improve optimization performance, a strategy with coarse-to-fine optimization is designed and outlined in Algorithm 2.

Algorithm 2 Non-Rigid Registration by the EAT Model

Input: the set of matching point pairs $F = \{(\mathbf{r}_k, \mathbf{s}_k)\}_{k=1}^K$, parameters σ_e and λ

Output: the optimal EAT matrix \mathbf{A}_e

- 1 Construct the set of the polynomial vectors for $\{\mathbf{r}_k\}_{k=1}^K$, which is represented as $\{\pi_k\}_{k=1}^K$;
- 2 Initialize the transformation parameter vector \mathbf{p} to zero vector, and set $w_2 = w_3 = w_4 = 2 \times 10^{-4}$, $w_5 = 0$;
- 3 By using the derivative (11), optimize the objective function (10) by the quasi-Newton method, and then obtain the optimal parameter vector \mathbf{p}_c of coarse optimization;
- 4 Initialize the transformation parameter vector to \mathbf{p}_c , and set $w_2 = w_3 = w_4 = 2 \times 10^{-4}$, $w_5 = 1 \times 10^{-7}$;
- 5 By using the derivative (11), re-optimize the objective function (10) by the quasi-Newton method, and then obtain the optimal parameter vector \mathbf{p}_o of fine optimization;
- 6 The optimal EAT matrix \mathbf{A}_e is computed by (9) with \mathbf{p}_o .

The coarse optimization implemented by the EAT model with mainly quartic polynomial model consists of the second and third steps of Algorithm 2. With the optimal result of the coarse optimization as initial value, the fine optimization consisting of the fourth and fifth steps is then achieved by the EAT model with mainly quintic polynomial model. The optimal EAT matrix \mathbf{A}_e indicates the regular pattern of non-rigid deformation between a pair of IR and VIS images.

To calculate each pixel's transformation, the set containing the polynomial vectors of all pixels in an IR image (it is represented as $\{\pi_q\}_{q=1}^Q$, where Q represents the number of pixels in an IR image) is constructed firstly and then,

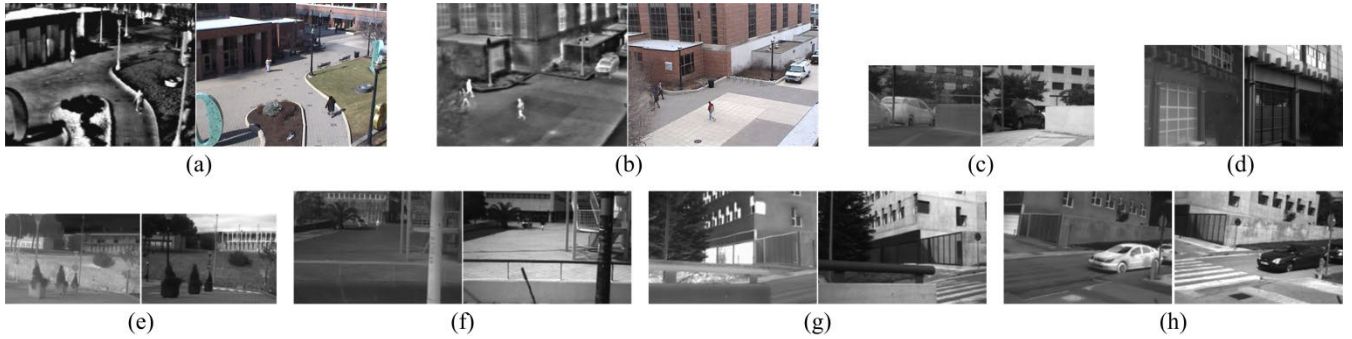


FIGURE 2. Dataset of IR and VIS images.

the corresponding transformation result is obtained by (8) with \mathbf{A}_e and $\{\pi_q\}_{q=1}^Q$. Because there may be some blank areas in an image after transformation, an image interpolation algorithm such as bilinear interpolation is required.

C. COMPUTATIONAL COMPLEXITY

As seen in the objective function IV-B and the corresponding derivative (11), their time complexity are both $O(K)$. In the quasi-Newton method, we use Armijo criteria to calculate the optimal search step and determine when to stop the optimization procedure. Thus, at each iteration, the calculation of the objective function needs to be performed twice and the calculation of the corresponding derivative needs to be performed once. Because the size of the EAT matrix \mathbf{A}_e is fixed, the total time complexity for solving optimization problem at each iteration is $O(K)$. It's worth mentioning that K is significantly less than the number of edge points in IR and VIS images. This is favorable for reducing the runtime of optimization.

The space complexity for solving the optimal EAT matrix is $O(K)$ due to the requirements of storing the set of the polynomial vectors $\{\pi_k\}_{k=1}^K$ with the size of $K \times 21$. Similarly, the space complexity for the spatial transformation of an IR image can be written as $O(Q)$. This is helpful to deal with large scale problems.

D. IMPLEMENTATION DETAILS

The state-of-the-art methods of non-rigid registration, such as TPS-RPM [33], CPD [38] and RGF [34], require data normalization so that it has zero means and unit covariance. However, floating point arithmetic can increase the difficulty of hardware implement, especially when dealing with the problem of precision and overflow. The integer coordinates of an image can be used for our method directly without any normalization. This can simplify the calculation steps and make our method easy to be implemented.

Our method requires the parameters to be set as follows: $\varepsilon_r = \varepsilon_v = \varepsilon_{rv} = 0.8$, $\sigma_e = 6$ and $\lambda = 0.02$. ε_r , ε_v and ε_{rv} are used to adjust the neighborhood sizes of GWSC. The range of Gaussian field in the objective function is defined by σ_e . λ is employed to regulate the trade-off between the closeness of two point sets and the smoothness of the solution. These parameters and the weight vector \mathbf{w} were determined through multiple experiments and kept constant throughout this work.

In addition, the influence of the parameter settings is shown in Section V. D.

V. EXPERIMENT

In this section, we tested the performance of the proposed GWSC at first. Our method was then compared with the state-of-the-art methods on real IR and VIS images. At last, an ablation study of the proposed EAT model is reported. The experiments were implemented by the computer with 3.9GHz Intel Core CPU, 4GB memory and Matlab code.

A. DATASET

In this work, the proposed method was evaluated on the dataset of real IR and VIS images. The dataset selected from OTCBVS [43] and CVC [18] datasets consists of the eight pairs of IR and VIS images, as shown in Fig. 2. The different intensity distributions and missing structures of the IR and VIS images are challenging for extraction of matching point pairs and estimation of transformation models. Meanwhile, it is obviously that there exists non-rigid deformation between the IR and VIS images. Therefore, by using this dataset, the performance of the proposed GWSC-EAT can be assessed effectively. In addition, the resolutions of the image pairs (a) and (b) are both 320×240 . The resolutions of (c), (d) and (e) are 189×136 , 166×170 and 227×151 , separately. The resolutions of (g) and (h) are both 283×189 .

To establish the ground truth, we manually selected actual matching point pairs in each image pair. The average number of actual matches selected manually in an image pair of our dataset is approximately 101.5. On the basis of the ground truth, quantitative evaluation can be implemented by using recall as the metric. For an actual matching point pair $(\mathbf{r}_g, \mathbf{s}_g)$, the mapping of \mathbf{s}_g in transformed images is represented as \mathbf{s}_g^t . If the Euclidean distance $\|\mathbf{r}_g - \mathbf{s}_g^t\|$ falls in an accuracy threshold (e. g., 3 pixels), we consider that \mathbf{s}_g is aligned to \mathbf{r}_g correctly. Therefore, the recall can be defined as the ratio between the number of actual matches aligned by registration method correctly and the number of actual matches in the ground truth.

B. EVALUATION OF GWSC

Although many improved algorithms of SC have been proposed, the most of them are applied to shape matching or point set registration. Since it is difficult to determine significant shape in the real images of the dataset (Fig. 2),

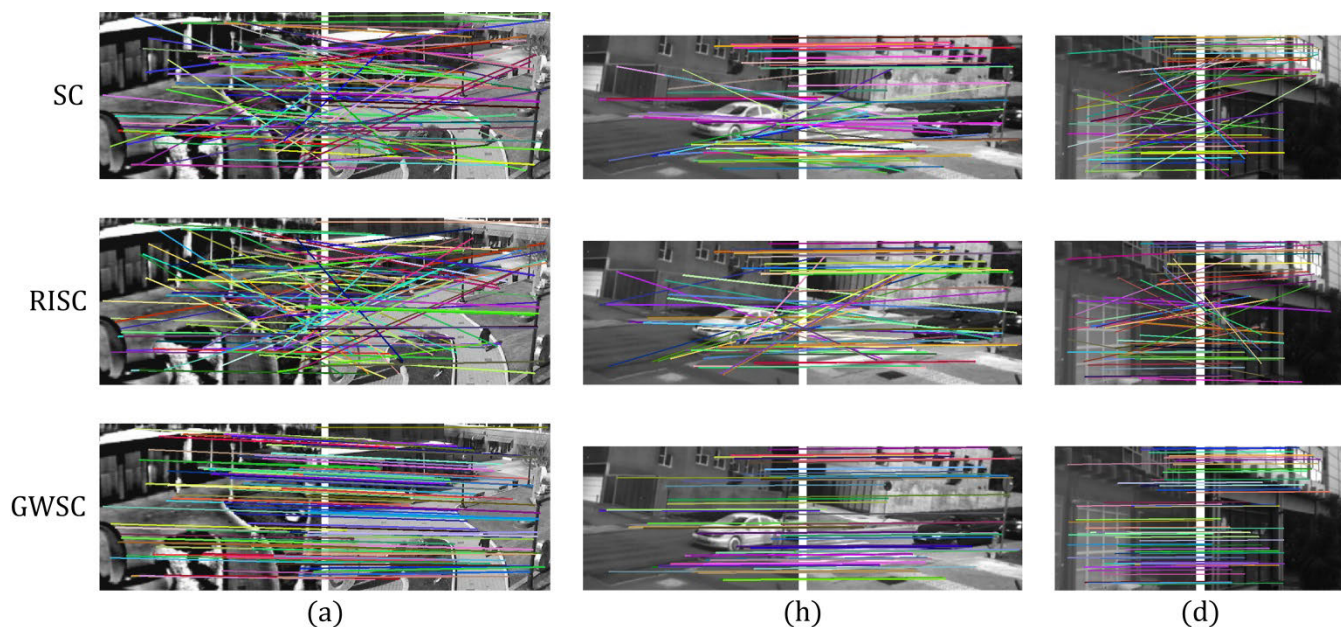


FIGURE 3. The illustrations of point matching by SC, RISC and GWSC on the dataset (a), (h) and (d). The lines indicate the correspondence pairs.

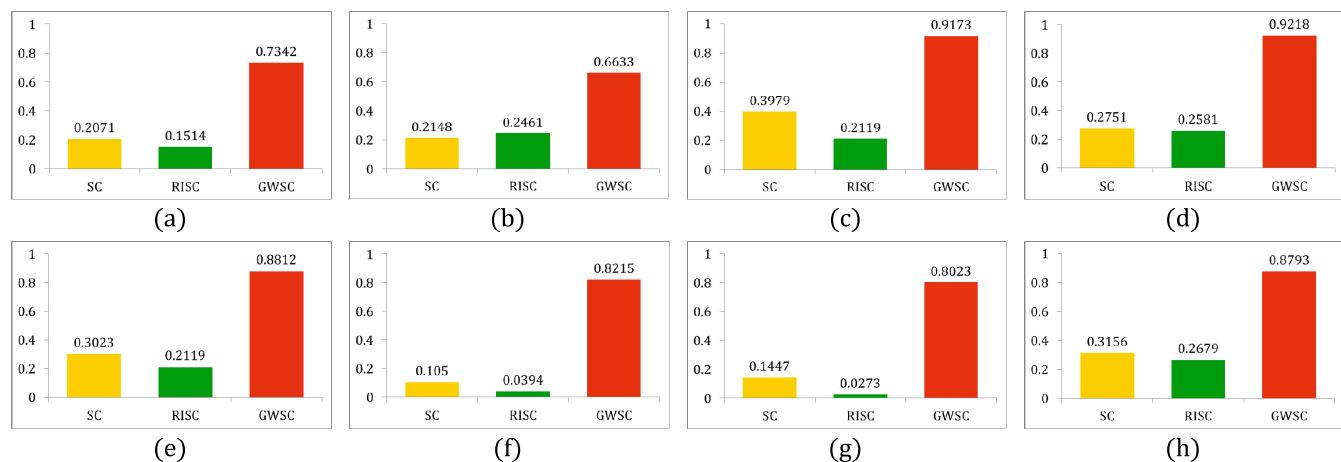


FIGURE 4. The recalls of point matching by SC, RISC and GWSC on the dataset.

some feature descriptors, such as IDSC [14], CDSC [15] and NWSC [16], cannot be used to point matching in IR and VIS images. Hence, in this section, the proposed GWSC was compared with the original SC [13] and RISC [17] on ground truth of the dataset. In order to test the performances of the feature descriptors, the bipartite graph matching of SC, RISC and GWSC was all implemented by the Hungarian method [39].

Fig. 3 shows the qualitative results of point matching by SC, RISC and GWSC on part of the dataset, where the lines indicate the correspondence pairs. Fig. 4 reports the quantitative results of point matching on the dataset. It can be seen that the qualitative results conform to the corresponding recalls well. The average recall of GWSC is 82.8%, while those of SC and RISC are 24.5% and 17.7%. On the one hand, this shows that RISC is not suitable for point matching in multi-modal images. On the other, the qualitative and quantitative comparisons demonstrate that the proposed GWSC is able to

significantly improve the accuracy on point correspondence of IR and VIS images.

The runtimes of point correspondence by SC, RISC and the proposed GWSC were also tested on ground truth. The computation times of the cost matrixes with SC, RISC and GWSC are all about 1.03 mins for 830 pairs of points. Actually, the major difference among the runtimes of point correspondence by the different approaches is the runtime of bipartite graph matching, as shown in Table 1. We can see that for 830 pairs of points in ground truth, the average runtime of bipartite graph matching with GWSC is about 0.47 mins, while those of SC and RISC are about 2.31 mins and 2.68 mins respectively. It is obviously that GWSC significantly improves the efficiency of bipartite graph matching. The reason for this is that the cost matrix with GWSC can describe the degree of correspondence between two point sets more accurately so that the number of augmenting paths in corresponding bipartite graph is less. Thus, this again

TABLE 1. The runtimes of bipartite graph matching with SC, RISC and GWSC on ground truth (min).

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
SC	5.2521	6.0665	0.1622	1.0077	1.2336	2.0191	1.1319	1.6165
RISC	8.0301	5.0174	0.1859	2.1930	2.1366	1.5593	1.1035	1.2182
GWSC	1.1155	0.7050	0.0207	0.3882	0.0701	1.0058	0.3512	0.1122

demonstrates that the proposed GWSC is more robust than SC and RISC for point matching in IR and VIS images.

There are mainly two differences between SC and the proposed GWSC. Firstly, it is assumed in GWSC that structure feature is similar between the points in the neighborhoods of two corresponding points. In other words, one point is determined to be matched with another point if all points in the neighborhoods of these two points exhibit similarity in terms of spatial structure. Secondly, it is assumed in GWSC that the displacement between one point in IR image and its corresponding point in VIS image is not over large. This is easy to be implemented by adjusting the optical axes of IR and VIS cameras in real applications. Based on the above two assumptions, non-corresponding points are more distinguishable. This is helpful to improve the robustness of point correspondence. The experiment results also show the effectiveness of the proposed GWSC.

C. COMPARISON EVALUATION WITH THE STATE-OF-THE-ART METHODS

In this section, the performance of the proposed GWSC-EAT was compared with those of the five state-of-the-art methods: CPD [38], RGF [21], RPM- L_2E [34], MR-RPM [17] and SC-TPS [13], which are able to estimate transformation model from point feature. To be fair, the point sets obtained by Algorithm 1 were used for all the above methods on an individual image pair. In this work, the MATLAB codes of these registration methods were provided by their authors and the parameters of these methods are the same as those set by their authors. To present registration results visually, an image fusion method with bilateral filter was employed in this experiment. The detail layer is extracted from a VIS image and fused with the corresponding IR image by an average fusion strategy.

The qualitative results of the various methods are shown in Fig. 5, including the results of edge map registration and image registration. Firstly, since all registration results are based on the feature point sets obtained by Algorithm 1, the qualitative results demonstrate that Algorithm 1 finds enough corresponding feature for accurate matching and performs well on IR and VIS image registration. Secondly, MR-RPM has excellent performance of point set registration, but it does not work on image registration because the transformation models estimated from local feature by MR-RPM are not able to achieve global image registration accurately. In general, the results of RPM- L_2E and RGF are better than those of CPD and SC-TPS. This demonstrates that the Gaussian field criterion is helpful for improving the accuracy of non-rigid registration. Thirdly, from the qualitative results

of Fig. V-D we can see that the registration quality of the proposed GWSC-EAT is greatly superior to those of CPD, RPM- L_2E , MR-RPM and SC-TPS, while it is slightly better than that of RGF. Next, the quantitative comparisons are required to further test the performance of the proposed method.

Fig. 6 reports the quantitative comparisons of CPD, RGF, RPM- L_2E , MR-RPM, SC-TPS and GWSC-EAT on the dataset. We can see that in most cases, the recall curves of our method are significantly above those of the other approaches when the accuracy thresholds are more than about 3 pixels. This demonstrates that the other methods may perform well on non-rigid registration for some local areas of images, but the proposed method is able to achieve more accurate global registration of IR and VIS images. This is also supported by the average matching errors of the various approaches on this dataset. The total average matching error of GWSC-EAT is about 2.27 pixels, while those of CPD, MR-RPM, RGF, RPM- L_2E and SC-TPS are about 3.97, 7.56, 2.97, 3.81 and 4.80 pixels. Compared with the other approaches, our method reduces the average matching error by about 50.8%.

It is obviously that the EAT model is the key element making our method better than the state-of-the-art methods. The EAT model is different from the existing non-linear transformation models: it does not require control points to achieve non-linear image transformation. In the existing non-linear transformation models, such as the TPS model, the RKHS-based model and the B-spline model, control points can be regarded as the key parameters of transformation models. However, the selection of control points is difficult to be optimized for image registration. The performance of transformation models may be limited with low quantity of control points, while computation complexity is significantly increased with large quantity of control points. Moreover, the location distribution of control points also influences the accuracy of image registration. Because the parameters of transformation models with control points are mainly optimized in the neighborhoods of control points, the mapping of every point is determined by nearby control points. As a result, the registration accuracy may be degraded in the points which are relatively far from control points. According to the above analyses, we find that the performances of transformation models with control points are limited by the uncertainty of control points. In other words, transformation models with control points rely heavily on local feature so that the precision of non-rigid image registration is limited.

There is none of the above issues in our method, because the EAT model consisting of the affine model and the polynomial model requires no control points. From (9) we can see that except for the weight vector \mathbf{w} , all parameters

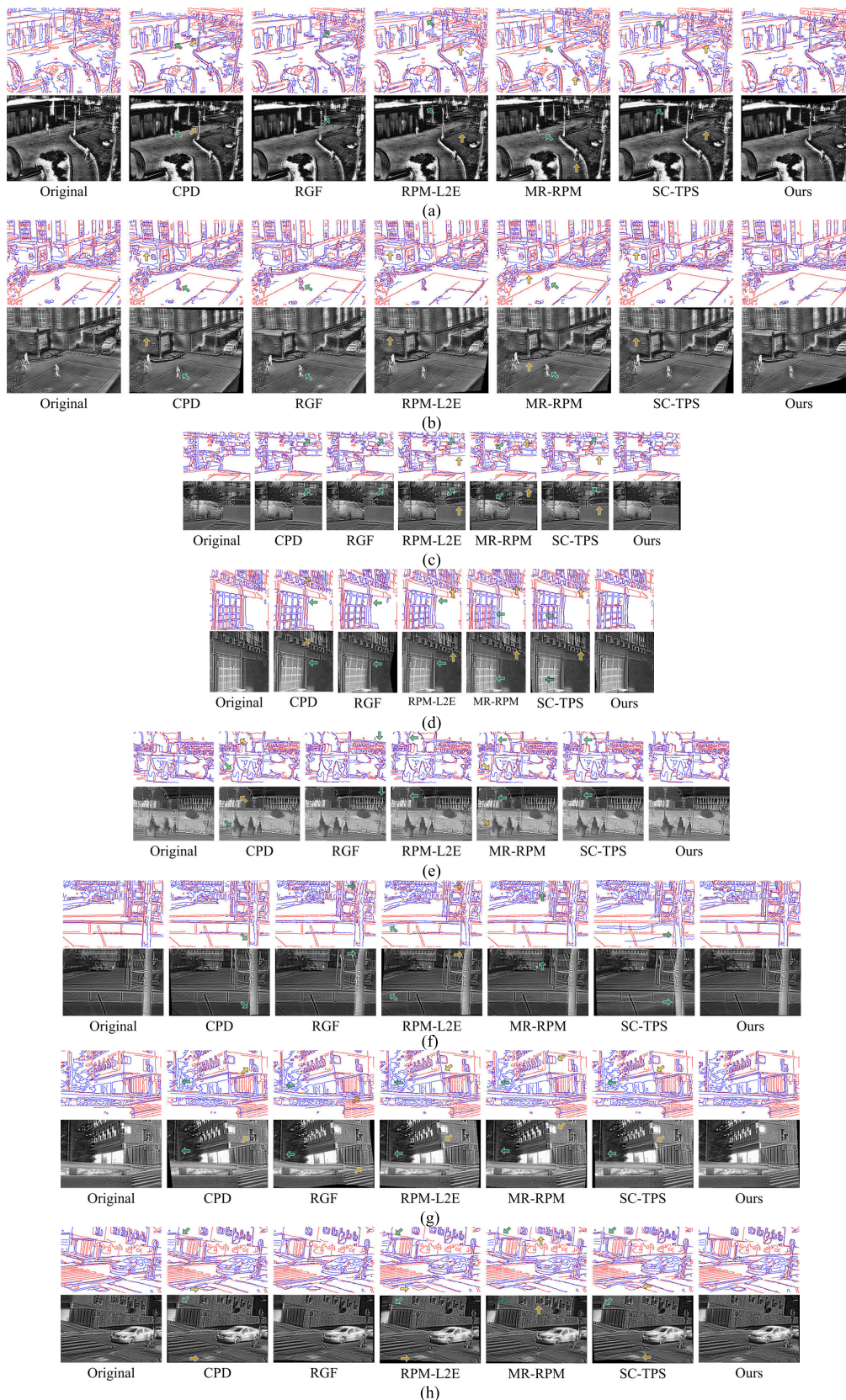


FIGURE 5. Qualitative registration results of CPD, RGF, RPM-L₂E, MR-RPM, SC-TPS and the proposed GWSC-EAT on the dataset. The blue and red lines separately represent the boundaries of IR and VIS images. The arrows highlight regions where there exist significant differences between the registration results of GWSC-EAT and the other approaches.

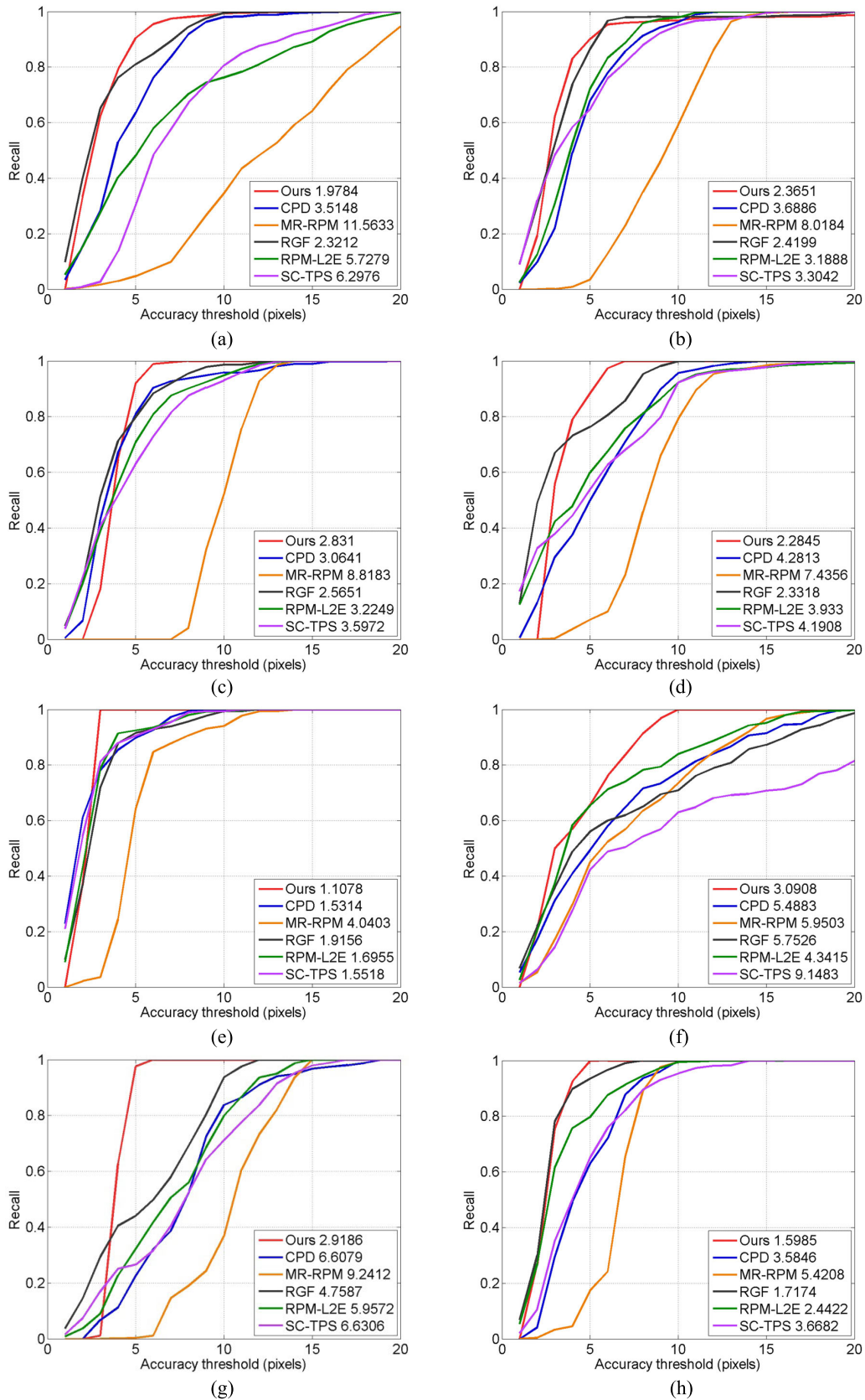


FIGURE 6. Quantitative comparisons of CPD, RGF, RPM-L₂E, MR-RPM, SC-TPS and the proposed GWSC-EAT on the dataset. The numbers in the legend are the average matching errors.

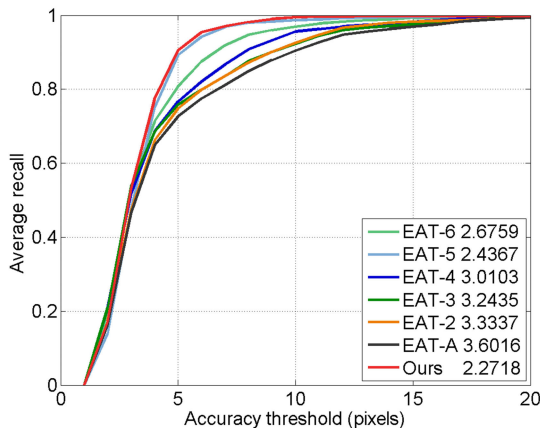


FIGURE 7. Quantitative comparisons of the EAT model with various orders and registration strategies on the dataset. The numbers in the legend are the total average matching errors.

of the EAT model can be optimized for image registration. This is a reason why our method is more accurate than the state-of-the-art methods. In addition, the EAT model prefers to describe the regularity of global deformation rather than local deformation nearby feature points. With the EAT model, the mapping of every point is determined by a regular pattern (8). Hence, although the EAT model is estimated from feature point sets, non-rigid image transformation can be implemented without feature point sets in our method. The distances between non-feature and feature points do not affect the spatial transformation with the EAT model in non-feature points. Thus, the EAT model is able to reduce the dependence of non-rigid image transformation on local feature and improve registration accuracy.

D. AN ABLATION STUDY OF THE PROPOSED EAT MODEL

In order to test the performance of the EAT model, we compared our method with the simplified versions of GWSC-EAT. Fig. 7 reports the total average recall curves and the total average matching errors of the corresponding results on the dataset. In the legend of Fig. 7, “Ours” represents the complete version of GWSC-EAT, while the others represent the simplified versions of our method. The simplified versions are modified from GWSC-EAT with the various weight vector \mathbf{w} and only coarse optimization.

EAT-A denotes the simplified version with $\mathbf{w} = 0$ (i.e. the affine model). EAT-2 denotes the simplified version with $w_2 = 2 \times 10^{-4}$ and $w_3 = w_4 = w_5 = 0$. EAT-3 denotes the simplified version with $w_2 = w_3 = 2 \times 10^{-4}$ and $w_4 = w_5 = 0$. EAT-4 denotes the simplified version with $w_2 = w_3 = w_4 = 2 \times 10^{-4}$ and $w_5 = 0$. EAT-5 denotes the simplified version with $w_2 = w_3 = w_4 = 2 \times 10^{-4}$ and $w_5 = 1 \times 10^{-7}$. EAT-6 denotes the simplified version with $G(x, y, \boldsymbol{\alpha}) = \sum_{i=2}^6 \sum_{j=0}^i w_i \alpha_{i,j} x^j y^{i-j}$ and $G(x, y, \boldsymbol{\beta}) = \sum_{i=2}^6 \sum_{j=0}^i w_i \beta_{i,j} x^j y^{i-j}$, where $w_2 = w_3 = w_4 = 2 \times 10^{-4}$, $w_5 = 1 \times 10^{-7}$ and $w_6 = 1 \times 10^{-10}$. Thus, the highest orders of the EAT models in EAT-2~EAT-6 are 2, 3, 4, 5 and 6 respectively.

Firstly, we can see in Fig. 7 that the non-linear transformation models are superior to the affine model (EAT-A) in

non-rigid registration of IR and VIS images. It demonstrates that the non-linear part in the proposed EAT model improves the performance of non-rigid registration.

Secondly, it is found that the higher order of the EAT model is not always better. The EAT models with the highest orders ≤ 5 are getting better with higher order, while the EAT model with the highest order = 6 is worse than EAT-5. This demonstrates two points: 1) the EAT model with low non-linearity is not enough to accurately describe the regular pattern of global deformation between IR and VIS images; 2) the EAT model with over high non-linearity can enlarge the influence of initial error between IR and VIS images, resulting in performance degradation of the quasi-Newton method. Thus, according to Fig. 7, we set the highest order of the EAT model to 5 in our method.

Thirdly, the experiment results show that the complete version of GWSC-EAT has better performance than the simplified versions. Particularly, although the highest order of the EAT model of our method is the same as that of EAT-5, our method is superior to EAT-5. It is proved that the strategy with coarse-to-fine optimization is able to weaken the impact of initial error and raise the chance of reaching global optimal.

The influence of the parameter settings was also investigated on our datasets. The results are shown in Fig. II. Obviously, it can be seen that the best performance of GWSC-EAT is achieved at $\varepsilon_r = \varepsilon_v = \varepsilon_{rv} = 0.8$, $\sigma_e = 6$, $\lambda = 0.02$, $w_2 = w_3 = w_4 = 2 \times 10^{-4}$ and $w_5 = 1 \times 10^{-7}$. In addition, When $w_2, w_3, w_4 > 2 \times 10^{-4}$ or $w_5 > 1 \times 10^{-7}$, the matching error is increased rapidly. This is because the EAT model with greater \mathbf{w} has more wide range so that the mappings of pixels are easy to exceed the size of images. Thus, it is proved that the weight vector \mathbf{w} is able to control the EAT model in a reasonable range.

E. RUNTIMES

We also tested the runtime of the proposed method and compared it with RGF and RPM- L_2E . These methods for image registration can be divided into three steps: extraction of feature points, estimation of transformation and image transformation. In real applications, for two point sets each containing 1000 boundary points, the average runtime of feature point extraction by the Hungarian method with GWSC is about 10 mins, while the fast feature point extraction (Algorithm 1) can reduce the runtime to about 1.3 mins. According to the above result of runtime and the qualitative results in Fig. 5, it is proved that without complex algorithms of bipartite graph matching, the proposed GWSC performs well on matching point extraction from IR and VIS images. This also demonstrates the advantage of GWSC on robustness.

Table 2 reports the runtimes of transformation estimation and image transformation, which are indicated as T_e and T_i respectively. The T_e of our method is significant lower than those of RGF and RPM- L_2E thanks to the simplified objective function (11). Meanwhile, the T_i of our method is slightly higher than that of RGF because our method has more

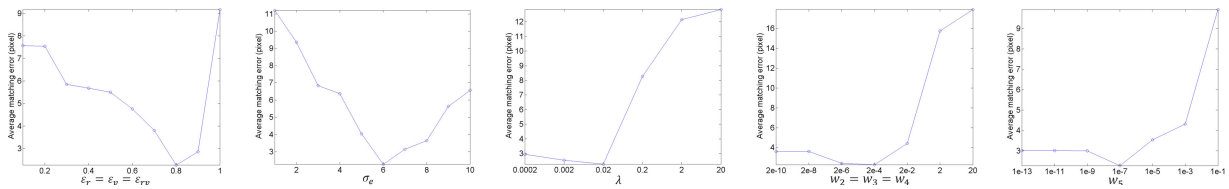


FIGURE 8. Illustration of the influence of parameter settings on the dataset.

TABLE 2. The total average runtimes of transformation estimation and image transformation (s).

	T_e	T_i	Total
RGF	33.5938	0.0441	33.6379
RPM- L_2E	19.9242	0.3591	20.2833
Ours	5.1337	0.0955	5.2292

parameters than RGF in this experiment. The dimension of the parameter vector of our method is 41×1 , while that of RGF with 15 control points is 15×2 in this work. From the total runtimes we can see that on the basis of the same feature point sets, our method is able to achieve fastest registration of IR and VIS images compared with RGF and RPM- L_2E . In addition, RGF is the typical method that solves the correspondence by optimization, while RPM- L_2E and our method both require potential correspondence. From Table 2 we can see that the methods with potential correspondence are faster on transformation estimation. This shows that potential correspondence can reduce the computation complexity of transformation estimation.

VI. CONCLUSION

Image registration is a prerequisite for a good fusion of IR and VIS images. In this work, we proposed a method GWSC-EAT for non-rigid registration of IR and VIS images. On the basis of SC, the GWSC is proposed to extract matching point pairs from IR and VIS images without any complex algorithms of bipartite graph matching. Meanwhile, the EAT model generalizing affine model from linear to non-linear case is able to accurately describe the regularity of global deformation between IR and VIS images. The various experiments show that GWSC-EAT is superior to the state-of-the-art methods on the accuracy of non-rigid image registration. In addition, it is also proved that the framework of non-rigid image registration with potential correspondence can improve the speed of image registration.

The weakness of the proposed method is that GWSC-EAT relies on image edge. If there are too much fuzzy edges, or there exist no significant edges, the proposed method may fail in image registration because the accuracy of the distance measure may be decreased with lower quantity of edge points. Although the EAT model describing the regularity of global deformation can reduce the influence of this weakness, it is still a common problem for feature-based registration methods. In addition, the canny edge detector adopted in our method is old and not robust to the intensity inhomogeneity and complex multiplicative noise in infrared images. Therefore, in future, we have two research focuses: 1) develop novel edge-preserving filtering of infrared images; 2) find multiplicative-noise-robust feature

to reduce the dependence of multimodal image registration on image edge.

REFERENCES

- [1] A. P. James and B. V. Dasarthy, "Medical image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 19, pp. 4–19, Sep. 2014, doi: 10.1016/j.inffus.2013.12.002.
- [2] Y. Zhou, K. Gao, Z. Dou, Z. Hua, and H. Wang, "Target-aware fusion of infrared and visible images," *IEEE Access*, vol. 6, pp. 79039–79049, 2018, doi: 10.1109/ACCESS.2018.2870393.
- [3] Z. Zhu, M. Zheng, G. Qi, D. Wang, and Y. Xiang, "A phase congruency and local laplacian energy based multi-modality medical image fusion method in NSCT domain," *IEEE Access*, vol. 7, pp. 20811–20824, 2019, doi: 10.1109/ACCESS.2019.2898111.
- [4] H. Li, S. Liu, Q. Duan, and W. Li, "Application of multi-sensor image fusion of Internet of Things in image processing," *IEEE Access*, vol. 6, pp. 50776–50787, 2018, doi: 10.1109/ACCESS.2018.2868227.
- [5] H. Wu, S. Zhao, J. Zhang, and C. Lu, "Remote sensing image sharpening by integrating multispectral image super-resolution and convolutional sparse representation fusion," *IEEE Access*, vol. 7, pp. 46562–46574, 2019, doi: 10.1109/ACCESS.2019.2908968.
- [6] Z. Zhang, H. Li, and G. Zhao, "Bionic algorithm for color fusion of infrared and low light level image based on rattlesnake bimodal cells," *IEEE Access*, vol. 6, pp. 68981–68988, 2018, doi: 10.1109/ACCESS.2018.2880845.
- [7] G. Piella, "A general framework for multiresolution image fusion: From pixels to regions," *Inf. Fus.*, vol. 4, no. 4, pp. 259–280, Dec. 2003, doi: 10.1016/s1566-2535(03)00046-0.
- [8] H. Sheng, "Medical image registration method based on tensor voting and harris corner point detection," *J. Med. Imag. Health Informat.*, vol. 8, no. 3, pp. 583–589, Mar. 2018, doi: 10.1166/jmhi.2018.2322.
- [9] Y. Xiang, F. Wang, and 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.*, vol. 56, no. 6, pp. 3078–3090, Jun. 2018, doi: 10.1109/TGRS.2018.2790483.
- [10] W. Cao, F. Lyu, Z. He, G. Cao, and Z. He, "Multimodal medical image registration based on feature spheres in geometric algebra," *IEEE Access*, vol. 6, pp. 21164–21172, 2018, doi: 10.1109/ACCESS.2018.2818403.
- [11] H. A. Rashwan, S. Chambon, P. Gurdjos, G. Morin, and V. Charvillat, "Using curvilinear features in focus for registering a single image to a 3D object," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4429–4443, Sep. 2019, doi: 10.1109/TIP.2019.2911484.
- [12] V. A. Zimmer, M. Á. G. Ballester, and G. Piella, "Multimodal image registration using laplacian commutators," *Inf. Fusion*, vol. 49, pp. 130–145, Sep. 2019, doi: 10.1016/j.inffus.2018.09.009.
- [13] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002, doi: 10.1109/34.993558.
- [14] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 286–299, Feb. 2007, doi: 10.1109/TPAMI.2007.41.
- [15] R.-X. Hu, W. Jia, D. Zhang, J. Gui, and L.-T. Song, "Hand shape recognition based on coherent distance shape contexts," *Pattern Recognit.*, vol. 45, no. 9, pp. 3348–3359, Sep. 2012, doi: 10.1016/j.patcog.2012.02.018.
- [16] J. Zhang, "Normalized weighted shape context and its application in feature-based matching," *Opt. Eng.*, vol. 47, no. 9, Sep. 2008, Art. no. 097201, doi: 10.1117/1.2977524.
- [17] J. Ma, J. Wu, J. Zhao, J. Jiang, H. Zhou, and Q. Z. Sheng, "Nonrigid point set registration with robust transformation learning under manifold regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3584–3597, Dec. 2019, doi: 10.1109/TNNLS.2018.2872528.
- [18] X. Liu, Y. Ai, B. Tian, and D. Cao, "Robust and fast registration of infrared and visible images for electro-optical pod," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1335–1344, Feb. 2019, doi: 10.1109/TIE.2018.2833051.

- [19] Y. Chen, J. Zhao, Q. Deng, and F. Duan, "3D craniofacial registration using thin-plate spline transform and cylindrical surface projection," *PLoS ONE*, vol. 12, no. 10, Oct. 2017, Art. no. e0185567, doi: [10.1371/journal.pone.0185567](https://doi.org/10.1371/journal.pone.0185567).
- [20] W. Sun, W. J. Niessen, and S. Klein, "Randomly perturbed B-Splines for nonrigid image registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1401–1413, Jul. 2017, doi: [10.1109/TPAMI.2016.2598344](https://doi.org/10.1109/TPAMI.2016.2598344).
- [21] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized Gaussian fields criterion," *Pattern Recognit.*, vol. 48, no. 3, pp. 772–784, Mar. 2015, doi: [10.1016/j.patcog.2014.09.005](https://doi.org/10.1016/j.patcog.2014.09.005).
- [22] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, Sep. 1997, doi: [10.1109/ICCV.1995.466930](https://doi.org/10.1109/ICCV.1995.466930).
- [23] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, Jan. 1999, doi: [10.1016/S0031-3203\(98\)00091-0](https://doi.org/10.1016/S0031-3203(98)00091-0).
- [24] M. Bansal and K. Daniilidis, "Joint spectral correspondence for disparate image matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2802–2809, doi: [10.1109/CVPR.2013.361](https://doi.org/10.1109/CVPR.2013.361).
- [25] J. Modersitzki and E. Haber, "Intensity gradient based registration and fusion of multi-modal images," *Methods Inf. Med.*, vol. 46, no. 3, pp. 292–299, Jan. 2018, doi: [10.1160/ME9046](https://doi.org/10.1160/ME9046).
- [26] L. Li, "Research on mutual information registration method based on gradient edge information," in *Proc. Int. Conf. Cyber Secur. Intell. Analytics*, 2019, pp. 1211–1216, doi: [10.1007/978-3-030-15235-2_161](https://doi.org/10.1007/978-3-030-15235-2_161).
- [27] Y. Chen and C. Lin, "PCA based regional mutual information for robust medical image registration," in *Proc. 8th Int. Symp. Neural Netw.*, 2011, pp. 355–362, doi: [10.1007/978-3-642-21111-9_40](https://doi.org/10.1007/978-3-642-21111-9_40).
- [28] G. Piella, "Diffusion maps for multimodal registration," *Sensors*, vol. 14, no. 6, pp. 10562–10577, Jun. 2014, doi: [10.3390/s140610562](https://doi.org/10.3390/s140610562).
- [29] C. Wachinger and N. Navab, "Entropy and Laplacian images: Structural representations for multi-modal registration," *Med. Image Anal.*, vol. 16, no. 1, pp. 1–17, Jan. 2012, doi: [10.1016/j.media.2011.03.001](https://doi.org/10.1016/j.media.2011.03.001).
- [30] M. Bansal and K. Daniilidis, "Joint spectral correspondence for disparate image matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2802–2809, doi: [10.1109/CVPR.2013.361](https://doi.org/10.1109/CVPR.2013.361).
- [31] H. Lombaert, J. Sporring, and K. Siddiqi, "Diffeomorphic spectral matching of cortical surfaces," in *Proc. Inf. Med. Imag.*, 2013, pp. 376–389, doi: [10.1007/978-3-642-38868-2_32](https://doi.org/10.1007/978-3-642-38868-2_32).
- [32] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, pp. 977–1000, Jun. 2003, doi: [10.1016/s0262-8856\(03\)00137-9](https://doi.org/10.1016/s0262-8856(03)00137-9).
- [33] H. Chuia and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Underst.*, vol. 89, pp. 114–141, Oct. 2002, doi: [10.1016/S1077-3142\(03\)00009-2](https://doi.org/10.1016/S1077-3142(03)00009-2).
- [34] J. Ma, W. Qiu, and J. Zhao, "Robust L_2E estimation of transformation for non-rigid registration," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1115–1129, Mar. 2015, doi: [10.1109/TSP.2014.2388434](https://doi.org/10.1109/TSP.2014.2388434).
- [35] J. Ma, J. Jiang, C. Liu, and Y. Li, "Feature guided Gaussian mixture model with semi-supervised EM and local geometric constraint for retinal image registration," *Inf. Sci.*, vol. 417, pp. 128–142, Nov. 2017, doi: [10.1016/j.ins.2017.07.010](https://doi.org/10.1016/j.ins.2017.07.010).
- [36] C. Yang, Y. Liu, X. Jiang, Z. Zhang, L. Wei, T. Lai, and R. Chen, "Non-rigid point set registration via adaptive weighted objective function," *IEEE Access*, vol. 6, pp. 75947–75960, 2018, doi: [10.1109/ACCESS.2018.2883689](https://doi.org/10.1109/ACCESS.2018.2883689).
- [37] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992, doi: [10.1109/34.121791](https://doi.org/10.1109/34.121791).
- [38] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010, doi: [10.1109/TPAMI.2010.46](https://doi.org/10.1109/TPAMI.2010.46).
- [39] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Chelmsford, MA, USA: Courier Corporation, 1982.
- [40] M. Elhenawy and H. Rakha, "A heuristic for rebalancing bike sharing systems based on a deferred acceptance algorithm," in *Proc. 5th IEEE Int. Conf. Models Technol. for Intell. Transp. Syst. (MT-ITS)*, Jun. 2017, pp. 188–193, doi: [10.1109/MTITS.2017.8005663](https://doi.org/10.1109/MTITS.2017.8005663).
- [41] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, pp. 325–340, Dec. 1987, doi: [10.1007/bf02278710](https://doi.org/10.1007/bf02278710).

- [42] J. Lee, H. Tang, and J. Park, "Energy efficient canny edge detector for advanced mobile vision applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 4, pp. 1037–1046, Apr. 2018, doi: [10.1109/TCSVT.2016.2640038](https://doi.org/10.1109/TCSVT.2016.2640038).
- [43] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Comput. Vis. Image Understand.*, vol. 106, nos. 2–3, pp. 162–182, May/Jun. 2007, doi: [10.1016/j.cviu.2006.06.010](https://doi.org/10.1016/j.cviu.2006.06.010).



CHAOBO MIN was born in Changzhou, Jiangsu, China, in 1987. He received the B.S. degree in applied physics from Nanjing Normal University, China, in 2009, and the Ph.D. degree in optical engineering from the Nanjing University of Science and Technology, China, in 2014.

From 2014 to 2017, he was a Vice Senior Engineer with North Night Vision Technology Company Ltd., Nanjing, China. Since 2017, he has been a Lecturer and a M.S. Supervisor with the College of Internet of Things Engineering, Hohai University, Changzhou. His research interests include machine vision, infrared imaging, multimodal image fusion, and image registration.



YAN GU was born in Ezhou, Hubei, China, in 1983. She received the B.S. and M.S. degrees in optical engineering from the Nanjing University of Science and Technology, China, in 2005 and 2009, respectively.

She is currently a Senior Engineer with North Night Vision Technology Company Ltd., Nanjing, China. Her research interests include electro-optical detection, UV imaging, and image processing.



FENG YANG was born in Nantong, Jiangsu, China, in 1988. He received the B.S. and M.S. degrees in optical engineering from the Nanjing University of Science and Technology, China, in 2011 and 2014, respectively.

Since 2019, he has been a Vice Senior Engineer with North Night Vision Technology Company, Ltd., Nanjing, China. His research interests include image processing systems and optoelectronic imaging systems.



YINGJIE LI was born in Suzhou, Jiangsu, China, in 1987. He received the B.S., M.S., and Ph.D. degrees in optical engineering from the Nanjing University of Science and Technology, China, in 2009, 2012, and 2016, respectively.

Since 2019, he has been a Senior Engineer with North Information Control Research Academy Group Company, Ltd., Nanjing, China. His research interests include optoelectronic imaging systems and image processing.



WENJUN LIAN was born in Taiyuan, Shanxi, China, in 1994. She received the B.S. degree in optoelectronic information science and engineering from the Beijing Institute of Technology, in 2017, and the M.Sc. degree in optics and photonics from Imperial College London, in 2018.

Since 2019, she has been working at North Information Control Research Academy Group Company Ltd., Nanjing, China. Her research interests include optoelectronic imaging systems and image processing.

• • •