

Point Set Registration with Global-local Correspondence and Transformation Estimation

Su Zhang^{1,2,3}, Yang Yang^{1,2,3,*}, Kun Yang^{1,2,*}, Yi Luo^{1,2}, Sim Heng Ong⁴

¹School of Information Science and Technology, Yunnan Normal University

²The Engineering Research Center of GIS Technology in Western China, Ministry of Education, China

³Laboratory of Pattern Recognition and Artificial Intelligence, Yunnan Normal University

⁴Department of Electrical and Computer Engineering, National University of Singapore

*Yang Yang and Kun Yang are joint corresponding authors.

*Email: {yyang-ynu, kmdeynu}@163.com

Abstract

We present a new point set registration method with global-local correspondence and transformation estimation (GL-CATE). The geometric structures of point sets are exploited by combining the global feature, the point-to-point Euclidean distance, with the local feature, the shape distance (SD) which is based on the histograms generated by an elliptical Gaussian soft count strategy. By using a bi-directional deterministic annealing scheme to directly control the searching ranges of the two features, the mixture-feature Gaussian mixture model (MGMM) is constructed to recover the correspondences of point sets. A new vector based structure constraint term is formulated to regularize the transformation. The accuracy of transformation updating is improved by constraining spatial structure at both global and local scales. An annealing scheme is applied to progressively decrease the strength of the regularization and to achieve the maximum overlap. Both of the aforementioned processes are incorporated in the EM algorithm, a unified optimization framework. We test the performances of our GL-CATE in contour registration, sequence images, real images, medical images, fingerprint images and remote sensing images, and compare with eight state-of-the-art methods where our GL-CATE shows favorable performances in most scenarios.

1. Introduction

Non-rigid point set registration is an essential problem which constantly draws interest in various fields such as remote sensing, medical image registration, template matching for hand-written characters and fingerprint identification. TPSRPM [5], RPM-LNS[34], CPD [22, 21], GMMREG [8], L2E-RPM [17, 15], GLMDTPS [32], MoAGREG

[30] and PR-GLS [18] are some of the outstanding works. In this section, we briefly review these methods, and then summarize from three aspects including: (i) energy optimization framework, (ii) fuzziness of correspondence and (iii) spatial constraint, followed by the outline of our GL-CATE.

The key idea of TPSRPM [5] is to assume that each source point corresponds to a weighted sum of the target points. The weights are taken from a corresponding matrix whose entries are proportional to a Gaussian function of the pairwise Euclidean distances between the source and target point sets. The transformation is updated using thin-plate splines (TPS) [3, 29]. Zheng et al. [34] proposed a robust point matching by preserving local neighborhood structures (RPM-LNS) for non-rigid shape registration based on the graph theory. RPM-LNS employs the shape context (SC) descriptor [2, 11] to initialize the graph matching, the optimal match between two graphs is the one that maximizes the number of matched edges. CPD [22, 21] takes the source and target point sets as the centroid of components and data, respectively, and the registration is interpreted to a maximum likelihood estimation problem. The elegant expectation maximization (EM) algorithm framework [6, 23] is employed for parametric estimation. The transformation is modeled using Gaussian radial basis function (GRBF) and the motion coherence theory [33] is used to regularize the displacement field between the point sets. GMMREG [8] extends the idea of registration from fitting Gaussian mixtures to data to aligning two Gaussian mixture models (GMM). And the L_2 distance is used to measure the discrepancy of two Gaussian mixtures instead of the log-likelihood function. RPM-L2E [17, 15] introduces the L_2 minimizing estimate (L_2E) [1, 28] which is a robust estimator in statistics to estimate the transformation. The source point set and the estimated corresponding point

which is obtained by matching the SC descriptor are represented by two multi-dimensional normal distributions to be fitted. GLMDTPS [32] presents a mixture-feature based correspondence estimation method named as global and local mixture distance (GLMD). The global distance is the point-to-point Euclidean distance, and the local distance is obtained by summing the squared Euclidean distance between the i^{th} neighboring points according to the index. These two distances are combined to form a GLMD based cost matrix for correspondences estimation. MoAGREG [30] uses asymmetric Gaussian distribution [10] to represent each point set, it updates correspondences and transformations under the framework of TPSRPM. PR-GLS [18] acquires a binary corresponding matrix by matching the SC descriptors, this matrix is then used to improve CPD through directly assigning the membership probabilities of GMMs to close to one, if matched, or to close to zero, otherwise.

TPSRPM employs robust point matching (RPM) algorithm for energy optimization, it essentially involves a dual update process embedded within an annealing scheme, which is quite similar to the EM algorithm adopted by CPD and PR-GLS. RPM-LNS proceed by finding the matches that maximizes the number of matched edges, while GLMDTPS optimizes the GLMD cost matrix using the Jonker-Volgenant algorithm [9, 20]. For GMMREG, RPM-L2E and MoAGREG, they optimize energy by minimizing the discrepancy of two distributions. Overall, TPSRPM, RPM-LNS, CPD, GMMREG, RPM-L2E, MoAGREG, the applicability of these methods is limited by the single feature based correspondence estimations. For example, if two target points share the same Euclidean distance to a source point, we obtain equal correspondences, although the local structures are probably totally different. And for robust shape context or graph features, however, is unfavored by the assumption that the corresponding points have similar neighborhood structures [8]. Moreover, the rotation invariant shape context used in RPM-LNS, RPM-L2E and PR-GLS can be greatly deteriorated since this property requires the center of mass to be stable. GLMDTPS and PR-GLS employ both the global and local features. However, since outliers are not modeled, GLMDTPS is sensitive to outliers. And PR-GLS actually divides the original single optimization process into two, which are the linear assign problem [24, 13] of the shape context and the EM algorithm. This inconsistency may cause a decrease in performance, especially, when handling large amount of points.

Based on the assumption that each source point corresponds to a weighted sum of the target points, TPSRPM and CPD estimate the one-to-many fuzzy correspondences by the Euclidean distance based probability, since the magnitude of the searching range parameter goes from large to small, it is a global-to-local registration strategy. At a very

large searching range, the estimated corresponding point set is essentially very close to the center of mass of the target point set, which leads the source point set to collapse at the beginning, and to expand as the searching range decreases, thus they require relatively more iterations. Particularly, the center of mass changes relatively slightly when the target point set is heavily rotated, which results in a bad initial pose and finally a large deviation result. On the other hand, enforcing a one-to-one correspondence using binary matrix [17, 15, 32] is vulnerable to the presence of noise and outliers.

Methods [22, 21, 17, 15] using GRBF constrain their spatial transformation with the motion coherence theory [33]. In TPS, a term in the form of the space integral of the square of the second order derivatives which reflects the prior knowledge is also included for the same purpose [5, 34, 32]. Intuitively, these regularizations discourage mappings which are too arbitrary by forcing points to move coherently at global scale. However, they produce position deviations of the rest of the points when one point is mismatched, and may also be undesirable when source points need to be moved in different directions to match their target points at the same time [32].

In this paper, we present a new point set registration method with global-local correspondence and transformation estimation (GL-CATE). The pairwise Euclidean distance and shape distance (SD) are used as the global and local features, respectively. The SD which is based on the histograms generated by an elliptical Gaussian soft count strategy can quantify similar neighborhood structures. By the help of the SD, the constructed mixture-feature Gaussian mixture model (MGMM) obtains a sufficiently good initial pose and reliable correspondence estimation. The searching ranges of the two features are directly controlled by a bi-directional deterministic annealing scheme, which interchanges the statuses of the two features and leads a local-to-global registration strategy. The EM algorithm, a unified optimization framework is used to estimate the parameters of the MGMM. In E-step, the posterior probability matrices are obtained by measuring the similarities of the designed mixture-feature using Bayes' rule. In M-step, we minimize the expectation of the negative objective function in reproducing kernel Hilbert space (RKHS). A new vector based structure constraint term is used to regularize the transformation which complements the global coherence. The accuracy of transformation estimation is improved by constraining spatial structure at both global and local scales. An annealing scheme is applied to progressively decrease the strength of the regularization and to achieve the maximum overlap. We test the performances of our GL-CATE in contour registration, sequence images, real images, medical images, fingerprint images and remote sensing images, and compare with eight state-of-the-art methods where our

GL-CATE shows favorable performances in most scenarios.

2. Method

We denote the set of M source points by $\mathbf{Y}_{M \times D} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M\}^T$, and a set of N target points by $\mathbf{X}_{N \times D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}^T$. The goal is to recover the unknown non-rigid transformation $\mathcal{T}(\mathbf{Y}, \phi)$ registering \mathbf{Y} to \mathbf{X} with maximum point-wise overlap, where ϕ is a set of parameters. Based on the reasonable assumption that points from one set are normally distributed around points belonging to the other set [18], aligning \mathbf{Y} onto \mathbf{X} is considered as fitting M Gaussian components to N data, where achieving the maximum point-wise overlap is therefore taken as minimizing the negative GMM log-likelihood function. Let \mathbf{y}_m be the centroid of the m^{th} component, \mathbf{x}_n the n^{th} data. The probability density function is obtained as:

$$p(\mathbf{x}_n) = (1 - \zeta) \sum_{m=1}^M \mathcal{C}_{mn} f(\mathbf{x}_n | \mathbf{y}_m) + \zeta \frac{1}{N}, \quad (1)$$

where \mathcal{C}_{mn} is non-negative quantity with $\sum_{m=1}^M \mathcal{C}_{mn} = 1$, which is called the component densities of the mixture. $\frac{1}{N}$ is an additional uniform distribution with a weighting parameter ζ , $0 \leq \zeta \leq 1$ for outlier dealing. Based on the EM algorithm for GMM based clustering, we compute the posterior probability (E-step) as:

$$p_{mn} = \frac{\mathcal{C}_{mn} f(\mathbf{x}_n | \mathbf{y}_m)}{\sum_{j=1}^M \mathcal{C}_{jn} f(\mathbf{x}_n | \mathbf{y}_j) + \zeta \frac{1}{N}}, \quad (2)$$

the matrix $\mathbf{P}_{M \times N}$ can be regarded as the correspondence matrix, whereby $\hat{\mathbf{X}} = \mathbf{P}\mathbf{X}$, the weighted sums of \mathbf{X} , or to be more specifically, the putative target set, is obtained. The new mixture model parameters are found by minimizing (M-step):

$$Q = - \sum_{n=1}^N \sum_{m=1}^{M+1} p_{mn} \log (\mathcal{C}_{mn} f(\mathbf{x}_n | \mathbf{y}_m)) + \mathcal{R}(\mathcal{T}), \quad (3)$$

where the first term is the expectation of the negative log-likelihood function of GMMs. The non-rigid transformation is formulated as $\mathcal{T}(\mathbf{Y}, \phi) = \mathbf{Y} + \mathcal{V}(\mathbf{Y})$, which is a kernel-based displacement function derived by using calculus of variation [21]. To prevent the ill-posed problem caused by \mathcal{T} which is not unique, the second term $\mathcal{R}(\mathcal{T})$ is therefore employed for regularization. The EM algorithm proceeds iteratively by alternating between E-step and M-step until convergence. In this paper, the density function $f(\mathbf{x}_n | \mathbf{y}_m)$ is specified based on a new finite mixture model, named mixture-feature Gaussian mixture model (MGMM). Meanwhile, the regularization term $\mathcal{R}(\mathcal{T})$ is formulated at both global and local scales.

2.1. Mixture-feature Gaussian Mixture Model

For our MGMM, the density function $f(\mathbf{x}_n | \mathbf{y}_m)$ is defined as:

$$f(\mathbf{x}_n | \mathbf{y}_m) = \frac{1}{2\pi\sigma\beta} \exp \left[- \left(\frac{\Delta^G}{2\sigma^2} + \frac{\Delta^L}{2\beta^2} \right) \right], \quad (4)$$

where Δ^G and Δ^L denote the similarity measures using global and local features, respectively. $\sigma^2 \in (0, 1)$ and $\beta^2 \in (0, 1)$ are covariances. Note that equal isotropic covariances $\sigma^2 \mathbf{I}$, $\beta^2 \mathbf{I}$ and component densities $\mathcal{C}_{mn} = \frac{1}{M}$ are applied for all MGMM components, where \mathbf{I} is a $D \times D$ identity matrix.

The underlying assumption of the density function (4) is the decomposition of the process for human to recognize and categorize objects. Supposing such process is based on the linear combination among features such as Euclidean distance and density, etc. The priority of certain feature may change during the process. For instance, one can easily categorize different letters according to the feature of shape at the very beginning, whereafter the accuracy can be further optimized by involving other features in. Inspired by these facts, the bi-directional deterministic annealing scheme is employed to gradually interchange the priorities of the global and local feature discrepancies during the registration, which is equivalent to enhancing the robustness of the MGMM by directly controlling the fuzziness of the correspondence [4]. We define the temperature parameter as $T = -\frac{\tau}{l}$, where τ is the current iteration number, and l is a constant. σ^2 and β^2 are obtained by $\sigma^2 = e^T$ and $\beta^2 = e^{\frac{1}{T}}$ in each iteration.

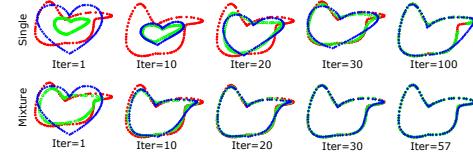


Fig. 1. The comparison on the registration processes of our GL-CATE using single feature (upper row) against its mixture-feature counterpart (lower row). Red asterisks: the target point set \mathbf{X} . Green circles: the estimated corresponding point set $\hat{\mathbf{X}}$. Blue crosses: the source point set \mathbf{Y} .

The advantage of MGMM can be demonstrated by comparing the registration processes of our GL-CATE using Δ^G only against its mixture-feature counterpart, as shown in Fig. 1. In this test, Δ^G is the point-to-point squared Euclidean distance, Δ^L is the squared shape distance which we will detail later. At the first iteration, we can see that by using Δ^G , the initial estimated coordinates (denoted by green circles) are the regional center of masses, while they are close to the real target coordinates (denoted by red asterisks) when using Δ^G and Δ^L . This helps us recover a good

initial pose for aligning the two point sets. At the 10th iteration, the estimated coordinates based on single feature discrepancy are still regional center of masses. By contrast, the counterpart estimates a preciser putative target by which a reliable source point set (denoted by blue pluses) is obtained. Finally, registration using single feature yields large point-to-point deviations, even if the iteration number is larger.

In the early stage of iterations, though the source point set \mathbf{Y} and the target point set \mathbf{X} have the biggest difference, the local feature can still be very strong and stable, and the correspondence which is estimated based on the local feature is therefore more reliable. This greatly facilitates the registration process at the very beginning. Other probability based methods (e.g., TPSRPM) can only yield a shrunken source point set near the regional center of masses. In our MGMM, since we have $\sigma^2 \approx 1$ and $\beta^2 \approx 0$, $\frac{\Delta^L}{2\beta^2}$ tends to be relatively large, thus the unreliable global correspondence is filtered due to the property of negative nature exponential function. At the final stage of iterations, \mathbf{Y} and \mathbf{X} are very similar, a direct estimation using the global feature is desirable. And the statuses of the two features interchange exactly since $\sigma^2 \approx 0$ and $\beta^2 \approx 1$, which means that the correspondence is mostly determined by the global feature.

2.2. MGMM based Correspondence Estimation

Typically, for point set registration, the original Mahalanobis distance which is a measure of distance between a point and a distribution is simplified to the Euclidean distance [4]. And in our study, we adopt the point-to-point Euclidean distance $\Delta_{mn}^G = \|\mathbf{x}_n - \mathcal{T}(\mathbf{y}_m, \phi)\|^2$ as the global feature. Based on the work in [2, 11], the shape distance

$$s_{mn} = \Delta_{mn}^L = \sum_{r=1}^{Rad} \sum_{t=1}^{Tan} \|c_n^{\mathbf{X}}(r, t) - c_m^{\mathbf{Y}}(r, t)\|^2 \quad (5)$$

is defined as the local feature, where $c(r, t)$ is the count of points within the rt^{th} bin, and Rad and Tan denote the number of bins in radial and tangential directions, respectively. However, when points lie close to the boundaries of bins, they may be assigned to different bins and yield biased histograms. And, non-rigid deformations occur in both radial and tangential directions, a soft count strategy which accordingly specifies the range and direction is more reasonable. Motivated by these facts, we present a new count strategy based on the elliptical Gaussian (EG) as:

$$c(r, t) = \begin{cases} EG(\mu^{rad}, \mu^{tan}, \Lambda^{rad}, \Lambda^{tan}), & \text{if } (r, t) \text{ are within } \mathcal{A} \text{ and } \mathcal{B} \\ 0, & \text{otherwise} \end{cases}, \quad (6)$$

where (μ^{rad}, μ^{tan}) is the coordinate of the EG centroid, \mathcal{B} is the polar coordinate with $Rad \times Tan$ bins. We respectively denote by ρ and θ the influence range in the radial

and tangential direction. The influence area \mathcal{A} centered at (μ^{rad}, μ^{tan}) is rectangle with lengths $2\rho + 1$ and $2\theta + 1$. Hence, the original binary count assigned at a single bin becomes a soft count assigned within \mathcal{A} . By choosing appropriate magnitudes for the two variances Λ^{rad} and Λ^{tan} , the EG weighting function can accurately quantify the non-rigid deformation. Table 1 shows the shape distances of two point sets using different count strategies. When $\theta = 0$, it is actually the original strategy that can not distinguish any of the five manners since the shape distances are all equal. The count by $\theta = 3$ distributes the best because the difference between adjacent columns are showing a progressively increasing tendency. This property makes the count be capable of identifying the similar neighborhood structures. A direct extension for 3D case can be realized by using 3DSC [7] and the ellipsoidal Gaussian soft technique, which spreads the counts along the azimuth, elevation and radial directions of the spherical coordinates. Since 2D case is the focus of this paper, we leave it to our future research.

Table 1. The shape distances of five manners using different count strategies.

$\theta = 0$	0	2	2	2	2
$\theta = 1$	0	1.0698	2.2707	2.5413	2.5413
$\theta = 2$	0	0.7058	2.2824	3.8210	4.6962
$\theta = 3$	0	0.5708	1.9058	3.6220	6.8896

Substituting the Euclidean distance and the elliptical Gaussian soft shape distance (EGSSD) into (4), we can therefore rewrite (4) in the complete form as:

$$f(\mathbf{x}_n | \mathbf{y}_m) = \frac{1}{2\pi\sigma\beta} \exp \left[- \left(\frac{\|\mathbf{x}_n - \mathcal{T}(\mathbf{y}_m, \phi)\|^2}{2\sigma^2} + \frac{s_{mn}}{2\beta^2} \right) \right]. \quad (7)$$

And the posterior probability function (2) can be rewritten as:

$$p_{mn} = \frac{\exp \left[- \left(\frac{\|\mathbf{x}_n - \mathcal{T}(\mathbf{y}_m, \phi)\|^2}{2\sigma^2} + \frac{s_{mn}}{2\beta^2} \right) \right]}{\sum_{j=1}^M \exp \left[- \left(\frac{\|\mathbf{x}_n - \mathcal{T}(\mathbf{y}_j, \phi)\|^2}{2\sigma^2} + \frac{s_{jn}}{2\beta^2} \right) \right] + g}, \quad (8)$$

where $g = \frac{2M\zeta\pi\sigma\beta}{N(1-\zeta)}$.

2.3. Global-local Structure Constraint (GLSC)

Once non-rigidity is allowed, there are an infinite number of ways to map one point set onto another. The ill-posed problem is prevented by term $\mathcal{R}(\mathcal{T})$ of objective (3). In our study, the regularization is defined as:

$$\mathcal{R}(\mathcal{T}) = \frac{\lambda}{2} \mathcal{G}(\mathcal{T}) + \frac{\eta}{2} \mathcal{L}(\mathcal{T}), \quad (9)$$

where parameters λ and η control the strength of regularizations, $\mathcal{G}(\mathcal{T}) = \|\mathcal{T}\|^2$ is the global structure constraint term following the motion coherence theory [33], and

$$\mathcal{L}(\mathcal{T}) = \sum_{m=1}^M \|\mathcal{E}(\hat{\mathbf{x}}_m) - \mathcal{E}(\mathcal{T}(\mathbf{y}_m, \phi))\|^2 \quad (10)$$

is the local structure constraint (LSC) term based on the local structure descriptor (LSD) $\mathcal{E}(\cdot)$. The extraction of the LSD is illustrated in Fig. 2. Given a set $\mathbf{Z}_{M \times D} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M\}^T$ of M points, let $\{\mathbf{z}_{ik}\}_{k=1}^K$ be the K nearest neighbors (KNN) of \mathbf{z}_i , $\{\mathbf{u}_{ik}\}_{k=1}^K$ the set with each entry denoting the vector $\overrightarrow{\mathbf{z}_i \mathbf{z}_{ik}}$, we have $\mathcal{E}(\mathbf{z}_i) = \sum_{k=1}^K h_{ik} \mathbf{u}_{ik}$, where $h_{ik} = \exp(-\|\mathbf{u}_{ik}\|^2/\nu_i^2)$ is the weight that controls the contribution of \mathbf{u}_{ik} to $\mathcal{E}(\mathbf{z}_i)$, and ν_i is the variance of $\{\|\mathbf{u}_{ik}\|\}_{k=1}^K$. $\mathcal{L}(\mathcal{T})$ exploits the local structural discrepancies between the putative target $\hat{\mathbf{X}}$ and the source $\mathcal{T}(\mathbf{Y}, \phi)$. Minimizing discrepancies $\mathcal{L}(\mathcal{T})$ is equivalent to leveraging the LSDs of $\mathcal{T}(\mathbf{Y}, \phi)$ and therefore forcing each neighboring point set of $\mathcal{T}(\mathbf{Y}, \phi)$ to align onto the corresponding one. A vivid demonstration is shown in Fig. 3.

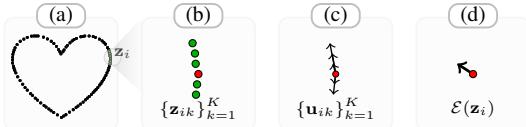


Fig. 2. Extraction of the local structure descriptor. (a): Selecting one point (colored in red), the goal is to obtain its LSD. (b): Finding its K nearest neighbors (colored in green). (c): Obtaining the respective vectors $\overrightarrow{\mathbf{z}_i \mathbf{z}_{ik}}$. (d): Computing the weighted vector sum. We set the weights h all equal to 1 and $K = 5$ for a brief demonstration. This figure is related to Fig. 3.

In addition, a deterministic annealing scheme is also applied to the trade-off parameters λ and η . The annealing parameter is defined as $\kappa = (\tau_{max}^4 - \tau^4 + 1)^{1/4}/\tau_{max}$ where τ_{max} is the maximum iteration number. The parameter κ remains stable in most iterations, then sharply decreasing to zero in the last several iterations. This implies that the constraint (9) is released at the final stage of iterations for achieving the maximization of the point-wise overlap.

2.4. GLSC based Transformation Estimation

Inspired by the Riesz representation theorem [27], we model the non-rigid transformation \mathcal{T} by lying it within the reproducing kernel Hilbert space (RKHS). We first define a RKHS \mathcal{H} by choosing a positive definite kernel, here we adopt the Gaussian radial basis function (GRBF). With the constant ϵ controlling the spatial smoothness, the kernel can be written in the form $\Theta(\mathbf{y}_i, \mathbf{y}_j) = \exp(-\frac{1}{2\epsilon^2} \|\mathbf{y}_i - \mathbf{y}_j\|^2)$

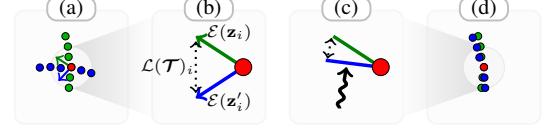


Fig. 3. The demonstration of how the LSC works. (a): Given a point set (colored in green) and its rotated form (colored in blue), the goal is to achieve maximum point-wise overlap between the k^{th} nearest neighbors of the center point (colored in red), as shown in (d). (b): Extracting the LSDs for the green and blue sets, the dotted black line denotes the local structural discrepancy. (c): Taking derivative of $\mathcal{L}(\mathcal{T})_i$ is equivalent to exerting an imaginary force on the LSD of the blue set. (d): The two sets are aligned. This figure is related to Fig. 2.

and easily generalized to three or higher dimensions. The optimal transformation function \mathcal{T} takes the form as:

$$\mathcal{T}(\mathbf{Y}, \mathbf{W}) = \mathbf{Y} + \Theta \mathbf{W}, \quad (11)$$

where $\mathbf{W}_{M \times D}$ is the coefficient matrix. Hence, the minimization over Q boils down to finding a finite coefficient matrix \mathbf{W} . The objective function (3) can be rewritten in a matrix form as:

$$Q(\mathbf{W}, \sigma^2, \beta^2) = \frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{m=1}^M p_{mn} \|\mathbf{X}_n - (\mathbf{Y} + \Theta \mathbf{W})_m\|^2 + \frac{\lambda}{2} \text{tr}(\mathbf{W}^T \Theta \mathbf{W}) + \frac{\eta}{2} \text{tr}(\mathbf{R} \mathbf{R}^T) + N_P \ln \sigma^2 \beta^2, \quad (12)$$

where $\mathbf{R} = (\mathbf{H} \hat{\mathbf{X}} - K \mathbf{I}) \hat{\mathbf{X}} - (\mathbf{H} \mathbf{Y} - K \mathbf{I})(\mathbf{Y} + \Theta \mathbf{W})$, \mathbf{I} is of size $M \times M$, and $N_P = \sum_{n=1}^N \sum_{m=1}^M p_{mn} \leq N$ (with $N_P = N$ only if $\zeta = 0$). Note that we omit term $\frac{1}{2\beta^2} \sum_{n=1}^N \sum_{m=1}^M p_{mn} s_{mn}$ in (12) since it is independent to \mathbf{W} . \mathbf{H}^Z is of $M \times M$ dimension with each non-zero entry $h_{ij} = \exp(-\|\overrightarrow{\mathbf{z}_i \mathbf{z}_j}\|^2/\nu_i^2)$, if $\mathbf{z}_j \in \{\mathbf{z}_{ik}\}_{k=1}^K$. Intuitively, the weighting matrix \mathbf{H}^Z is analogous to a permutation matrix, by which the weighted LSDs of point set \mathbf{Z} can be summed if multiplying \mathbf{H}^Z by \mathbf{Z} .

Taking partial derivative of (12) with respect to \mathbf{W} , we obtain the coefficient matrix \mathbf{W} as:

$$\mathbf{W} = (d(\mathbf{P} \mathbf{1}) \Theta + \lambda \sigma^2 \mathbf{I} + \eta \sigma^2 \mathbf{B}^T \mathbf{B} \Theta)^{-1} (\mathbf{P} \mathbf{X} - d(\mathbf{P} \mathbf{1}) \mathbf{Y} + \eta \sigma^2 \mathbf{B}^T \mathbf{A}), \quad (13)$$

where $\mathbf{A} = (\mathbf{H} \hat{\mathbf{X}} - K \mathbf{I}) \hat{\mathbf{X}} - (\mathbf{H} \mathbf{Y} - K \mathbf{I}) \mathbf{Y}$, $\mathbf{B} = \mathbf{H} \mathbf{Y} - K \mathbf{I}$ and $d(\cdot)$ denotes the diagonal of a matrix. So far all the parameters have been solved. The new location of the source point set is updated by (11), after which we return to correspondence estimation and continue the registration process until the maximum iteration number τ_{max} is reached.

3. Parametric Setting and Complexity Analysis

The expressions of the point sets on the coordinate system affect the performance of point set registration methods. Thus an initial normalization process is involved in our GL-CATE to rescale the two point sets \mathbf{Y} and \mathbf{X} to have zero means and unit variances. In each iteration, reasonable magnitude for β^2 is obtained by normalizing the counts to have unit length as $\sum_{r=1}^{Rad} \sum_{t=1}^{Tan} c(r, t)^2 = 1$, normalization $p_{mn} \leftarrow p_{mn} / \sum_{j=1}^M p_{mj}$ is also required for the computation of the estimated corresponding point set $\hat{\mathbf{X}} = \mathbf{P}\mathbf{X}$.

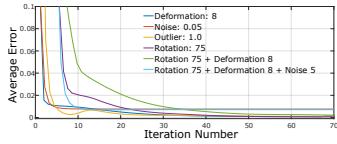


Fig. 4. Convergence experiment on synthesized dataset under the largest degree of six degradation categories.

Experiments show that our GL-CATE will catch a good stable solution after about 60 iterations, as shown in Fig. 4. We set $\tau_{max} = 60$ and $l = 5$. Thus $\sigma_{init}^2 = e^{-1/5} = 0.8187$, $\sigma_{final}^2 = e^{-60/5} \approx 6.144 \times 10^{-6}$, $\beta_{init}^2 = e^{-5/1} = 0.0067$, $\beta_{final}^2 = e^{-5/60} \approx 0.9200$. We find that when the target point set is contaminated by noise, the performance can be improved by computing σ^2 as:

$$\sigma^2 = \frac{tr(\mathbf{X}^T d(\mathbf{P}^T \mathbf{1}) \mathbf{X}) - 2tr(\mathbf{V}^T \mathbf{P} \mathbf{X}) + tr(\mathbf{V}^T d(\mathbf{P} \mathbf{1}) \mathbf{V})}{2N_P} \quad (14)$$

instead of using $\sigma^2 = e^T$, where $\mathbf{V} = \mathcal{T}(\mathbf{Y}, \mathbf{W})$. In noise scenario, precise alignment is not preferred, and registration process using (14) is similar to the linear least-squares fitting, which makes more sense. We set $\zeta = 0.7$ for noise and outlier scenarios, otherwise $\zeta = 0.3$, and covariances σ^2 are computed by (14) in all noise contaminated scenarios.

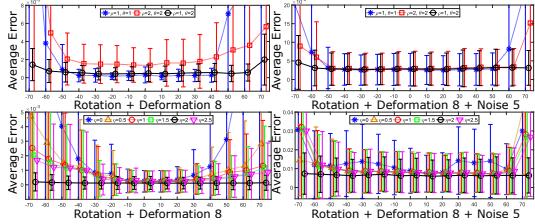


Fig. 5. Experiments on synthesized dataset using different EG soft count strategy (upper row) and LSC smoothness (lower row). Left column: RD8 scenario. Right column: RD8N5 scenario.

Parameters ρ and θ for elliptical Gaussian (EG) weighting function, and parameter η for the smoothness of the local structure constraint (LSC) are tested under two extreme scenarios, as shown in Fig. 5. The polar coordinate \mathcal{B} is set to 5×10 , each bin in tangential direction stands for

36° . Variances of EG function are set as $\Lambda^{rad} = \rho^2$ and $\Lambda^{tan} = \theta^2$. When the degree of rotation is smaller than 40° , setting $\rho = 1$, $\theta = 1$ performs better than setting $\rho = 1$, $\theta = 2$, while the opposite occurs under larger degree. Therefore, we set $\rho = 1$ and $\theta = 2$ for scenarios with rotation, and both to 1 for others. The number of the nearest neighboring points K is set to 5 to distinguish between a corner (includes two neighboring points) and a cross (includes four neighboring points) [32]. The regularization parameters λ and η are both set to 2. ϵ determines the width of the range of the interaction between samples, we empirically set ϵ to 2. The pseudo-code of our GL-CATE is shown in Algorithm 1.

Algorithm 1: The GL-CATE Method

```

input : Two point sets  $\mathbf{Y}$  and  $\mathbf{X}$ 
output : Transformed point set  $\mathbf{Y}$ 

1 Construct the kernel matrix  $\Theta$ ;
2 Initialization:  $\mathbf{W} = \mathbf{0}$ ;  $\lambda = \eta = 2$ ;  $\zeta = 0.7$  under
   noise and outlier, otherwise  $\zeta = 0.3$ ;  $K = 5$ ;
3 Compute the histograms of  $\mathbf{X}$  by (6);
4 while not convergence do
5   E-Step:
6     Compute the histograms of  $\mathbf{Y}$  by (6);
7     Compute the shape distance between  $\mathbf{X}$  and
        $\mathbf{Y}$  by (5);
8     Compute the posterior probability matrix  $\mathbf{P}$ 
       by (8);
9     Compute temperature parameter  $T = -\frac{\tau}{l}$ ;
10    Compute  $\kappa = \frac{(\tau_{max}^4 - \tau^4 + 1)^{1/4}}{\tau_{max}}$ ;
11    M-Step:
12      Compute  $\mathbf{W}$  by (13);
13      Update the source point set  $\mathbf{Y} = \mathbf{Y} + \Theta \mathbf{W}$ ;
14      Update covariances  $\sigma^2$  by (14) under noise,
         otherwise by  $\sigma^2 = e^T$ ;
15      Update covariances  $\beta^2 = e^{\frac{1}{T}}$ ;
16      Anneal  $\lambda = \kappa \lambda$  and  $\eta = \kappa \eta$ ;
17 end
18 The transformed source point set  $\mathbf{Y}$  is obtained in
   the final iteration.

```

For our GL-CATE, it takes $\mathcal{O}((CM)^2)$ to obtain the shape distances for M points by using the soft count strategy, where $C = (2\theta + 1)(2\rho + 1)$. The K nearest neighbors are obtained by K operations of sequential search with $\mathcal{O}(KM)$ complexity for one point. The heat kernel \mathbf{H} requires $\mathcal{O}(KM^2)$ to compute. The derivative (13) is of $\mathcal{O}(M^3)$ due to the existence of $M \times M$ kernel Θ . Overall, the computational complexity of GL-CATE is $\mathcal{O}(M^3)$, which, as well, can be ulteriorly reduced to $\mathcal{O}(M)$ complexity by using several well-studied techniques such as low-rank matrix approximation [19] and subset of regres-

sors method [26, 16]. For other methods, TPSRPM, CPD, GMMREG and MoAGREG require $\mathcal{O}(MN)$ work to compute the corresponding matrix, where $M \leq N$. Instead, to solve the linear assignment problem based on the $N \times N$ dummy cost matrix, RPM-L2E, GLMDTPS and PR-GLS need $\mathcal{O}(N^3)$ work by which a permutation matrix is obtained. The regular complexity for the radial basis function (e.g., thin-plate spline and Gaussian radial basis function) based transformation estimation step is $\mathcal{O}(M^3)$, and for CPD, RPM-L2E, MoAGREG and PR-GLS, the fast implementation which has $\mathcal{O}(M)$ complexity is presented in their literatures.

4. Experiments

Both the basic and application experiments are carried out to compare the performance of our GL-CATE against eight state-of-the-art methods which are TPSRPM[5], RPM-LNS[34], CPD[21], GMMREG[8], RPM-L2E[17], GLMDTPS[32], MoAGREG[30] and PR-GLS[18]. The experiments are implemented in Matlab on a laptop with 2.60 GHz Intel Core CPU, 16 GB RAM.

4.1. Results on Basic Experiments

In the first series of experiments, four synthesized contour data sets, *fish1*, *Chinese character*, *hand* and *line*, each of which respectively contains 98, 105, 302 and 60 points are used. Four degradation categories, i.e., deformation, noise, outlier and rotation are used to evaluate the accuracy and robustness of methods. We follow the experimental settings as in [32], which are degree of deformation from 1 to 8, noise level from 0.01 to 0.05, outlier to data ratio from 0.2 to 1.0, and enlarging the rotation range from $\pm 30^\circ$ to $\pm 90^\circ$, with a 15° interval. We abbreviate the n^{th} degree as Dn , Nn , On and Rn for convenience of description. Note that $D4$ degradation is contained in noise, outlier and rotation scenarios by default. The average performances on the four data sets are shown in the first row of Fig. 6. Our GL-CATE gives the best performances over $D1$ to $D8$, $N1$ to $N5$ and $O4$ to $O5$. In rotation scenario, it yields near-perfect result within $\pm 75^\circ$ rotation.

In the second series of experiments, four synthesized contour data sets, *butterfly*, *face*, *fish2* and *maple*, each of which respectively contains 172, 317, 98 and 215 points are used. Each data is first degraded by either $D8$ or $D4 + N5$, and then rotated by -80° to 80° with a 10° interval. We abbreviate as $RD8$ and $RD4N5$, respectively. We also test the average runtime of our GL-CATE on all the six scenarios, using three data sets with representative point numbers. The average performances of methods using four data sets on $RD8$ and $RD4N5$, as well as the average runtime are shown in the second row of Fig. 6. We see that our GL-CATE outperforms within $\pm 70^\circ$ rotation in $RD8$ and $RD4N5$ scenarios. The results of RPM-LNS, RPM-L2E

and PR-GLS using rotation invariant shape context are deteriorated since the mass centers of the data are changed in such cases. Formally, our GL-CATE takes about 1 second to register two point sets with $M, N = 100$ points.

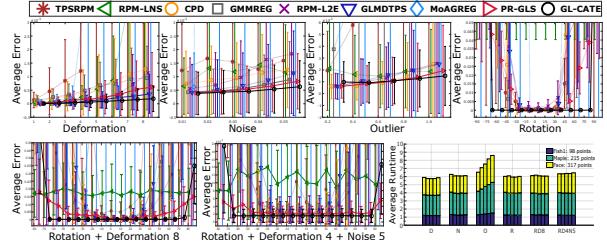


Fig. 6. Comparison of our GL-CATE against eight state-of-the-art methods on eight point sets and showcase of the average runtime of our GL-CATE. The error bars indicate the standard deviations of the average errors in 100 random experiments. In each line graph, methods with representative performances are highlighted. The stacked bar graph is used to demonstrate the average runtime, for each bar group, five degrees ranged from the first to the last are selected for showcase, and the units are in seconds.

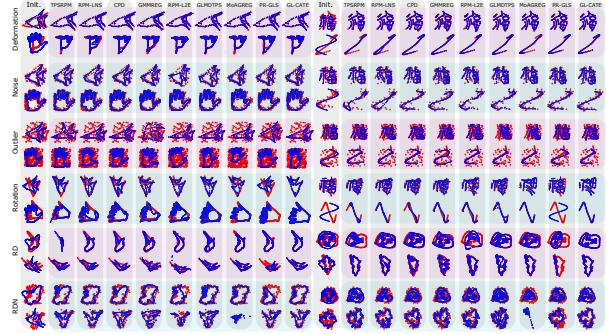


Fig. 7. Registration examples on *fish1*, *Chinese character*, *hand*, *line*, *butterfly*, *face*, *fish2* and *maple*, with deformation, noise, outliers, rotation, RD and RDN scenarios being shown in every two rows. The goal is to align the source point set (blue asterisks) onto the target point set (red circles). All the six scenarios are at the following degrees: (i) degree of deformation 8; (ii) noise level 0.05; (iii) outlier to data ratio 1.0; (iv) degree of rotation 75° ; (v) degree of rotation $75^\circ +$ degree of deformation 8 and (vi) degree of rotation $75^\circ +$ degree of deformation 4 + noise level 0.05.

In the third series of experiments, three types of data sets are used. (i) A dataset from [12] which consists of 30 pairs of car images and 20 pairs of motorbike images selected from Pascal 2007 Challenge, each pair contains 30–60 feature points. (ii) CMU hotel which consist of 101 frames and each frame has 30 labeled landmarks, all possible image pairs are tested. (iii) IMM which consists of face landmarks with expression and multi-view changes, each face contains 58 point landmarks with different facial expressions and poses, 8 groups of face data are tested. The matching rate of these three experiments are listed in Table 2A.

Table 2. Quantitative comparisons on multiple experiments. A: The matching rate on cars, motorbikes, CMU hotel and IMM images. B,C and D: The means and standard deviations of RMSE, MAE and MEE on retinal images, fingerprint images and remote sensing images, respectively. \times denotes that the number of the points is less than the minimum requirement of RPM-LNS. Bold values indicate the best performances.

	TPSRPM	RPM-LNS	CPD	GMMREG	RPM-L2E	GLMDTPS	MoAGREG	PR-GLS	GL-CATE
A	Car	80.67%	\times	82.83%	78.30%	81.69%	87.07%	77.45%	92.06%
	Motorbike	84.49%	\times	84.25%	83.29%	83.45%	99.40%	74.74%	90.77%
	Hotel	94.69%	89.02%	98.42%	97.14%	98.06%	90.77%	99.81%	99.16%
	IMM	80.62%	77.35%	82.74%	80.50%	91.68%	93.05%	70.43%	79.02%
B	RMSE	5.24 ± 3.50	\times	4.45 ± 2.61	4.73 ± 3.84	4.49 ± 0.15	4.38 ± 0.92	5.45 ± 2.50	4.46 ± 0.64
	MAE	6.30 ± 4.00	\times	5.66 ± 2.79	6.03 ± 4.87	5.92 ± 0.51	5.35 ± 1.37	6.64 ± 2.61	5.34 ± 0.63
	MEE	2.36 ± 0.99	\times	1.92 ± 1.41	2.17 ± 1.81	2.51 ± 0.81	1.84 ± 0.75	2.09 ± 0.94	1.89 ± 0.41
C	RMSE	14.57 ± 10.10	15.98 ± 21.67	20.72 ± 26.93	16.97 ± 18.44	15.99 ± 4.92	15.60 ± 12.83	10.26 ± 9.89	7.15 ± 5.67
	MAE	18.46 ± 12.62	19.29 ± 25.42	26.04 ± 34.94	21.00 ± 21.82	19.48 ± 5.72	17.17 ± 16.22	12.87 ± 13.36	9.48 ± 7.85
	MEE	2.33 ± 0.46	3.80 ± 2.67	4.75 ± 8.25	4.32 ± 4.67	3.87 ± 2.83	2.50 ± 6.30	3.40 ± 1.83	2.03 ± 1.61
D	RMSE	4.06 ± 4.71	3.63 ± 5.37	1.96 ± 1.09	5.71 ± 8.23	9.96 ± 17.64	3.76 ± 4.85	2.24 ± 1.03	3.49 ± 7.88
	MAE	5.03 ± 5.88	4.52 ± 6.63	2.38 ± 1.25	7.20 ± 9.86	13.00 ± 23.20	4.70 ± 6.28	2.86 ± 1.50	4.30 ± 9.88
	MEE	1.45 ± 1.63	1.53 ± 1.26	1.10 ± 1.24	4.08 ± 7.11	7.35 ± 13.11	1.75 ± 3.01	1.79 ± 4.00	1.33 ± 0.56

4.2. Results on Application Experiments

In the fourth series of experiments, three types of data sets are used. (i) A pair of CT images and two pairs of MRI images from [25], the images have a resolution in the range from 515×460 to 620×458 . The point sets are created by the edges of the objects and extracted using a regional scanning method. The recall curve is used to evaluate performances under 40 landmarks from edges. The results are shown in Fig. 8. (ii) Ten pairs of retinal images in multi-view scenarios, the images have a resolution in the range from 640×480 to 1280×960 . (iii) Eight pairs of fingerprint images from BIT (<http://biometrics.idealtest.org/dbDetailForUser.do?id=7>), each image is of resolution 640×480 pixels. The feature points of (ii) and (iii) are extracted following the method proposed in [31]. (iv) Five pairs of remote sensing images (Venice, Hawaii, London, New York and Dali) from Google Earth, the images have a resolution in the range from 800×600 to 1150×700 . For experiments on (ii)–(iv), they are all registered using SIFT feature points [14]. At least ten pairs of corresponding points between the transformed and reference images are manually determined for quantitative comparisons based on the root of mean square error (RMSE), maximum error (MAE) and median error (MEE), all corresponding points are well-distributed and selected on the easily identified areas, and the results are listed in Table 2B–D, respectively. Registration examples of (ii)–(iv) are shown in Fig. 9.

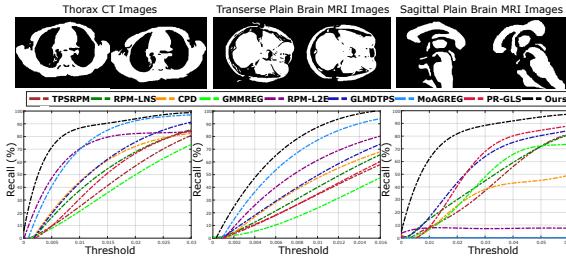


Fig. 8. Registration performances on CT and MRI images.

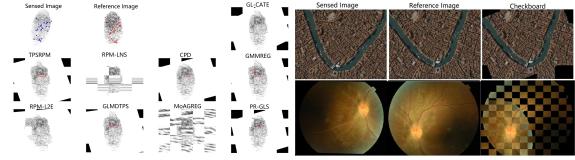


Fig. 9. Examples on fingerprint images (left), remote sensing images (top right) and retinal images (bottom right).

5. Conclusion

We have presented a dual-feature based point set registration method with global-local spatial constraint. The main idea of our GL-CATE is to first estimate the correspondence using the mixture-feature Gaussian mixture model (MGMM), and then to update the transformation under global-local spatial constraint. Comparing with the current methods, the major contributions of this work are: (i) the MGMM which can deal with two features for estimating correspondence is constructed, meanwhile, a uniform distribution is modeled for outlier dealing; (ii) a new elliptical Gaussian soft histogram count strategy as well as the derived shape distance is presented, which is able to quantify similar neighborhood structures; (iii) a bi-directional deterministic annealing scheme is used to combine the Euclidean distance and the shape distance with the MGMM, which leads to a local-to-global registration strategy; (iv) a new Tikhonov regularization term is designed, this term and the original one play complementary roles to improve the robustness and accuracy of transformation at both global and local scales. Experimental results demonstrate that the proposed GL-CATE achieve better performances than state-of-the-art methods.

6. Acknowledgment

This work was supported by (i) National Nature Science Foundation of China [41661080]; (ii) Scientific Research Foundation of Yunnan Provincial Department of Education [2017TYS045].

References

- [1] A. Basu, I. Harris, N. Hjort, and M. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85:549–559, 1998.
- [2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:509–521, 2002.
- [3] F. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11:567–585, 1989.
- [4] H. Chui and A. Rangarajan. A feature registration framework using mixture models. In *Proc. IEEE MMBIA-2000*, 190–197, Hilton Head Island, SC, 2000.
- [5] H. Chui and A. Rangarajan. A new algorithm for non-rigid point matching. *Comput. Vis. Image Understand.*, 89:114–141, 2003.
- [6] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. Series B*, 39:1–38, 1977.
- [7] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *Proc. Eur. Conf. Comput. Vis.*, Pittsburgh, PA, May 2004.
- [8] B. Jian and B. Vemuri. Robust point set registration using gaussian mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33:1633–1645, 2011.
- [9] R. Jonker and A. Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38:325–340, 1987.
- [10] T. Kato, S. Omachi, and H. Aso. Asymmetric Gaussian and its application to pattern recognition. *Structural, Syntactic, and Statistical Pattern Recognition*, 405–413, Springer, Berlin, Heidelberg, 2002.
- [11] M. Körtgen, M. Novotni, and R. Klein. 3d shape matching with shape context. In *The 7th Central European Seminar on Computer Graphics*, 2003.
- [12] M. Leordeanu, R. Sukthankar, and M. Hebert. Unsupervised learning for graph matching. *Int. J. Comput. Vis.*, 96:28–45, 2012.
- [13] H. Ling and D. Jacobs. Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:286–299, 2007.
- [14] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60:91–110, 2004.
- [15] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. Yuille, and Z. Tu. Robust L₂E estimation of transformation for non-rigid registration. *IEEE Trans. Signal Proc.*, 63:1115–1129, 2015.
- [16] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu. Regularized vector field learning with sparse approximation for mismatch removal. *Pattern Recognit.*, 46:3519–3532, 2013.
- [17] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. Yuille. Robust estimation of nonrigid transformation for point set registration. *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2147–2154, 2013.
- [18] J. Ma, J. Zhao, and A. Yuille. Non-rigid point set registration by preserving global and local structures. *IEEE Trans. Image Process.*, 25:53–64, 2016.
- [19] I. Markovsky. Low rank approximation. Springer, Verlag London, 2011.
- [20] M. Miller, H. Stone, and I. Cox. Optimizing murty’s ranked assignment method. *IEEE Trans. Aerosp. and Electron. Syst.*, 33:851–862, 1997.
- [21] A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32:2262–2275, 2010.
- [22] A. Myronenko, X. Song, and M. Carreira-Perpinan. Non rigid point set registration: Coherent point drift. *Advances in Neural Information Processing Systems*, 1009–1016, 2006.
- [23] M. Neal, Radford and G. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. *Learning in Graphical Models*, 355–368. Springer Netherlands, Dordrecht, 1998.
- [24] C. Papadimitriou and K. Steiglitz. Combinatorial optimization: Algorithms and complexity. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1982.
- [25] L. Peng, G. Li, M. Xiao, and L. Xie. Robust cpd algorithm for non-rigid point set registration based on structure information. *PLoS ONE*, 11(2):e0148483, 2016.
- [26] R. Rifkin, G. Ye, and T. Poggio. Regularized least-squares classification, *NATO science series: Computer and Systems Sciences*, 190:131–154. IOS Press, Amsterdam, The Netherlands, 2003.
- [27] B. Schölkopf, R. Herbrich, and A. Smola. A generalized representer theorem. In *International Conference on Computational Learning Theory*, 416–426, London, UK, 2001. Springer-Verlag.
- [28] D. Scott. Parametric statistical modeling by minimum integrated square error. *Technometrics*, 43:274–285, 2001.
- [29] G. Wahba. Spline models for observational data. *SIAM*, Philadelphia, Pennsylvania, 1990.
- [30] G. Wang, Z. Wang, Y. Chen, and W. Zhao. A robust non-rigid point set registration method based on asymmetric gaussian representation. *Comput. Vis. Image Understand.*, 141:67–80, 2015.
- [31] G. Wang, Z. Wang, Y. Chen, and W. Zhao. Robust point matching method for multimodal retinal image registration. *Biomed. Signal Process. Control*, 19:68–76, 2015.
- [32] Y. Yang, S. Ong, and K. Foong. A robust global and local mixture distance based non-rigid point set registration. *Pattern Recognit.*, 48:156–173, 2015.
- [33] A. Yuille and N. Grzywacz. A mathematical analysis of the motion coherence theory. *Int. J. Comput. Vis.*, 3:155–175, 1989.
- [34] Y. Zheng and D. Doermann. Robust point matching for nonrigid shapes by preserving local neighborhood structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28:643–649, 2006.