

Nonrigid Points Alignment with Soft-weighted Selection*

XueLong Li¹, Jian Yang¹, Qi Wang^{1,2*}

¹School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xian 710072, Shaanxi, P. R. China

²Unmanned System Research Institute (USRI), Northwestern Polytechnical University, Xian 710072, Shaanxi, P. R. China
xuelong_li@opt.ac.cn, jynwpu@mail.nwpu.edu.cn, crabwq@nwpu.edu.cn

Abstract

Point set registration (PSR) is a crucial problem in computer vision and pattern recognition. Existing PSR methods cannot align point sets robustly due to degradations, such as deformation, noise, occlusion, outlier, and multi-view changes. In this paper, we present a self-selected regularized Gaussian fields criterion for nonrigid point matching. Unlike most existing methods, we formulate the registration problem as a sparse approximation task with low rank constraint in reproducing kernel Hilbert space (RKHS). A self-selected mechanism is used to dynamically assign real-valued label for each point in an accuracy-aware weighting manner, which makes the model focus more on the reliable points in position. Based on the label, an equivalent matching number optimization is embedded into the non-rigid criterion to enhance the reliability of the approximation. Experimental results show that the proposed method can achieve a better result in both registration accuracy and correct matches compared to state-of-the-art approaches.

1 Introduction

As a crucial technique of computer vision, point set registration, is the process of recovering the correspondences and determining the spatial transformation between the point sets that contain the same context sampled at different time, from different viewpoints or by different sensors. In most applications, the matching results greatly influence the accuracy of the subsequent processing tasks like dynamic target detection, simultaneous localization and mapping (SLAM)[Hong and Kim, 2016] and object reconstruction [Zhou *et al.*, 2015]. Therefore, it is necessary to develop a robust method that can automatically align point sets with sub-pixel accuracy.

Recently, lots of PSR methods have been proposed, which mainly include two categories. One uses rigid transformation to model the spatial deformation between two input point sets, which is simple but has a lower accuracy for complex dis-

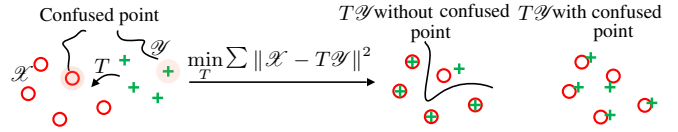


Figure 1: Example of PSR with confused point. The goal of PSR is moving data points \mathcal{Y} onto model points \mathcal{X} by the underlying spatial transformation T .

tortions, and the other models the spatial deformation as non-rigid transformation by a set of base functions. Since spatial deformation between point sets has multi-degree of freedom in real scenario, non-rigid registration is closer to real deformation. So, it is more widely used than the rigid ones. Owing to following difficulties, PSR becomes pretty challenging: (1) degradations: the distribution of point set would be more complex in the presence of large degradations, e.g., noise, occlusion, outlier, scale, and multi-view changes. The noisy data often means the points cannot be matched precisely. Moreover, the data with occlusion and outliers could cause some points fail to find the true correspondences. Consequently, the registration accuracy is unstable for these factors; (2) the numerical optimization often falls into local minima; (3) the computational complexity is higher when aligning the point set with great number.

For these challenges, earlier methods mainly applied distance ratio to recover point correspondences by comparing the distance of the closest neighbor with that of the second-closest neighbor [Lowe, 2004]. However, some outliers or missing matches always exist in this manner. Particularly, the outliers or missing matches become more serious as the distractions, e.g., noise, similar structure or improper threshold, increase. Note that the outliers have a negative effect on the estimation of spatial transformation while less matches mean more unreliable parameter calculation. To remove outliers, geometric structures [Zhang *et al.*, 2014] and spatial constraints [Kahaki *et al.*, 2016], are often used. Although these methods perform well, it is infeasible to align non-rigid point sets for the unknown underlying transformation.

Unlike above matching, another often-used approach is the one based on iterative optimization. Such methods first formulate point matching as an optimization problem, and then

*Corresponding author: crabwq@nwpu.edu.cn

solve it via numerical algorithms. The typical methods based on Gaussian fields criterion [Myronenko and Song, 2010] used a Gaussian mixture model to maximize the overlap between the features in both images. Whereas this criterion is generally differentiable, some gradient-based optimization algorithms can be directly utilized to solve it. Besides, it does not need explicit point correspondences and is more robust to noise as well. However, most of them are only suitable for rigid deformations. For the nonrigid case, Ma *et al.* [Ma *et al.*, 2015c] improved it via a regularized constraint and sparse approximation. Wang *et al.* [Wang *et al.*, 2017] used a laplacian regularized term to preserve the intrinsic geometry of the transformed set. Qu *et al.* [Qu *et al.*, 2017] solved nonrigid point matching problem by regression and clustering under a Bayesian framework. However, these methods only aim to minimize the global error. Since the position accuracies of different points are usually different, this means they should have different contributions to the accumulation of global error. In this sense, only minimizing the global error could make parameter estimation imprecise.

There are often some points (called confused samples) that should be outliers or noises in the transformed set, and the global error would be smaller when using them than without them as shown in Fig. 1. Such confused points easily mislead the iterative model, even make it stuck in local extremum. In this paper, we introduce a self-selected mechanism to remove the confused points by guiding the convergence of the model and propose a robust point matching criterion. Our main contributions are:

- We formulate the point matching problem as a sparse approximation with low rank constraint in RKHS, which makes our model more efficient.
- An adaptive self-selected mechanism is presented to remove the confused points, which adaptively select reliable points as iteration and make our the model focus more on the faithful samples;
- An equivalent matching number constraint is embedded into the non-rigid criterion, which enhances the reliability of point matching.

2 Related Work

Point matching has been widely applied to computer vision. Over past few years, many point matching methods have been proposed, which can be divided into two aspects. The first is the ones based on RANSAC [Lowe, 2004]. As a robust estimator, RANSAC tries to find a feasible subset to estimate the transformation by multiple sampling. Based on it, several improvements are presented [Chou and Wang, 2015]. For instance, Litman *et al.* [Litman *et al.*, 2015] proposed an inverting RANSAC to detect inliers in the given set. To get more matches, Wu *et al.* [Wu *et al.*, 2015] first applied a fast sample consensus (FSC) to obtain initial correct matches, and then tried to find the precise matches from the initial matches as many as possible. Although these methods perform well in most cases, they are designed for rigid matching and unsuitable for nonrigid alignment.

Another popular strategy is to iteratively match the points and solve the optimal spatial transformation, such as coher-

ence point drift (CPD) [Myronenko and Song, 2010], progressive vector field consensus (PVFC) [Ma *et al.*, 2015a], robust point matching with L_2E estimator (PRM- L_2E) [Ma *et al.*, 2015c], image gradient orientations (IGO) [Zheng *et al.*, 2017]. Besides, Tsin [Tsin and Kanade, 2004] employed the kernel correlation (KC) to align the points. Afterwards, Bing [Jian and Vemuri, 2011] introduced the Gaussian mixture models (GMMs) and expanded the method [Tsin and Kanade, 2004]. To preserve local or global structures better, Zheng and Doermann [Zheng and Doermann, 2006] constructed a simple graph matching initialized with the shape context distance, and proposed a robust point matching by preserving the local structures (RPM-LNS). Ma *et al.* [Ma *et al.*, 2016] formulated point alignment as the estimation of a mixture of densities. Bai *et al.* [Bai *et al.*, 2018] introduced local connectivity constraint by k -connected neighbors and connectivity matrix. In these works, PSR is often formulated as the optimal problems based on GMMs or graph matching [Pinheiro *et al.*, 2017], which has a limited calculation efficiency in real applications.

3 Methodology

3.1 Problem Formulation

Given two point sets, the model set $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^n$ and the data set $\mathcal{Y} = \{\mathbf{y}_r\}_{r=1}^m$, the goal of registration is to recover the correspondence between $\{\mathbf{x}_i\}_{i=1}^n$ and $\{\mathbf{y}_r\}_{r=1}^m$. This process is moving the model set \mathcal{X} onto the data set \mathcal{Y} by a series of transformation τ . In general, the best spatial transformation $\tau(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}^d$ is just the one that has the maximum point-to-point overlapping between \mathcal{X} and \mathcal{Y} by τ , where d is the dimension of one point. When we assume that the noise on the inlier correspondence $(\mathbf{x}_i, \mathbf{y}_r)$ is Gaussian on each component with zero mean and uniform standard deviation σ , the correspondence $(\mathbf{x}_i, \mathbf{y}_r)$ satisfies $\mathbf{y}_r - \tau(\mathbf{x}_i) \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, where \mathbf{I} is a $d \times d$ identity matrix. Thus, the point correspondence problem can be regarded as finding the largest samples (i.e., the inlier points) that obey the normal density model.

Defining displacement function $\mathbf{f}(\mathbf{x}) = \tau(\mathbf{x}) - \mathbf{x}$, above matching problem can be further written as follows

$$\min_{\mathbf{f}} - \sum_{i=1}^n \sum_{r=1}^m w_{ir} \exp(\eta_{ir}) + \lambda \Phi(\mathbf{f}), \quad (1)$$

where

$$\eta_{ir} = - \frac{d(\mathbf{x}_i + \mathbf{f}(\mathbf{x}_i), \mathbf{y}_r)^2}{\sigma^2}, \quad (2)$$

$\Phi(\mathbf{f})$ is a regularization term controlled by constant λ for trading-off the two terms, and the stabilizer $\Phi(\mathbf{f})$ is an inner product $\Phi(\mathbf{f}) = \|\mathbf{f}\|_{\Gamma}^2 = \langle \mathbf{f}, \mathbf{f} \rangle_{\Gamma}$, where $\|\cdot\|_{\Gamma}^2$ is the Hilbertian norm of the vector-valued RKHS, $d(\mathbf{x}, \mathbf{y})$ denotes the Euclidean distance between vector \mathbf{x} and \mathbf{y} , and w_{ir} is the weight coefficient.

In Eq. 1, Gaussian distance is used as measure function. And if model point \mathbf{x}_i and data point \mathbf{y}_r are matched, the Gaussian distance is equal to 1 for $d(\tau(\mathbf{x}_i), \mathbf{y}_r) = 0$, whereas if $\mathbf{x}_i, \mathbf{y}_r$ are not matched or outliers, $d(\tau(\mathbf{x}_i), \mathbf{y}_r)$ is generally larger and the Gaussian distance is closer to 0. So, Eq. 1 aims to find the largest matches. However, due to the disturbances

in the data, e.g., noises, local similar structures and minor outliers, the model is more easily confused so that converges to local extremum. To avoid above problems, a self-selected mechanism is introduced to guide model convergence by dynamic sample selection and incoordinate treatment in this paper. Specifically, sample label is assigned for each model point to select suitable sample points first. A self-selected function w.r.t. the label is defined to measure how much contribution each sample point devotes to the model, which can restrain the matching number and work for dynamically updating the label. As a result, we formulate the point alignment as the following self-selected regularized criterion:

$$\begin{aligned} \min_{\mathbf{f}, \mathbf{v}} \quad & \sum_{i=1}^n v_i \sum_{r=1}^m w_{ir} \ell(\gamma_{ir}) + \sum_{i=1}^n \kappa(v_i, \varsigma) + \lambda \Phi(\mathbf{f}), \\ \text{s.t.} \quad & \forall i, r, \gamma_{ir} = \frac{\|\mathbf{x}_i + \mathbf{f}(\mathbf{x}_i) - \mathbf{y}_r\|^2}{\sigma^2}; W \geq 0; v \in [0, 1]^n, \end{aligned} \quad (3)$$

where W is the weight matrix, written as $W = [w_1, \dots, w_m] \in \mathbb{R}^{n \times m}$ with $w_r = [w_{r1}, \dots, w_{rn}]^T$; $v_i \in [0, 1]$ is the sample label for the model point \mathbf{x}_i that shows whether it is selected. $\kappa(\cdot, \varsigma)$ is a function controlled by the parameter $\varsigma > 0$, which specifies how \mathbf{x}_i is assigned. $\ell(x) = 1 - \exp\{-x\}$ and $\|\cdot\|$ denotes the L_2 norm.

3.2 Self-selected Mechanism

Here, label v_i is used as a measure of selection or deselection, which can be performed by a hard-value within $\{0, 1\}$, i.e., setting selection with $v_i = 1$ while deselection with $v_i = 0$. But such hard-value cannot reflect the importance of different samples to the model. In this paper, we treat the selected samples with unequal weights. $v_i \in [0, 1]^n$ is set in a mixture of a hard 0-1 and a soft-assignment manner. More importantly, each v_i is dynamically updated in order to select the best sample points for current model during each iteration. For this purpose, we introduce a self-selected function $\kappa(v_i; \varsigma)$ w.r.t. v_i , just as in [Jiang *et al.*, 2014], and it is crystallized as follows

$$\kappa(v_i; \varsigma, \zeta) = -\zeta \ln(v_i + \zeta/\varsigma), \quad \varsigma, \zeta > 0, \quad (4)$$

where the extra parameter ζ is used to control the strength of the weights assigned to the selected points.

Let $\ell_i = \sum_r w_{ir} \ell(\rho_{ir})$ be the accumulated error of model point \mathbf{x}_i . To discuss v_i better, omit the items without v_i , and then for each model point \mathbf{x}_i , Eq. 3 can be simplified as

$$v_i^* = \arg \min_{v_i} v_i \ell_i + \kappa(v_i; \varsigma), \quad \text{s.t. } v_i \in [0, 1], \quad (5)$$

Note that Eq. 5 is continuous and differentiable with respect to v_i . Thus, it is easy to derive its optimal solution v_i^* by taking a derivative with respect to v_i , then we have

$$v_i^* = \begin{cases} 1, & \ell_i \leq \zeta\varsigma/(\zeta + \varsigma) \\ 0, & \ell_i \geq \zeta \\ \zeta/\ell_i - \zeta/\varsigma, & \text{otherwise,} \end{cases} \quad (6)$$

It is clear that each v_i is related to its accumulated error in Eq. 6. Smaller accumulated error means more contributions

to the model for the sample point with larger v_i , whereas the contribution of the sample point varies from small to nothing as ℓ_i increasing. Moreover, when ℓ_i is less than a small constant, the corresponding $v_i \in [0, 1]$ becomes 1, which enhances the influence of the less but faithful sample points. Conversely, $v_i \in [0, 1]$ is assigned as 0 and it works for reducing the effect of those unstable samples on the model. In other words, the accumulated error actually reflects the sample accuracy in position, which can be used to measure its importance to the model. Generally, higher accuracy means the point is more important and the sample point would be more suitable for the model. From this view, such strategy can select the faithful points in each iteration. As iteration, it makes our model preferentially select some reliable points and then gradually increase the points with larger influence on the model.

From Eq.6, each label v_i is updated according to ℓ_i , which makes our model focus more on the reliable and accurate points in each iteration. By iterative point selection, our method gradually selects the points from reliable ones to less reliable ones, which are continually added into the final point set. The final points can be directly applied to estimate the displacement function $\mathbf{f}(\mathbf{x})$ better and more reliably. In other words, this is an accuracy-aware matching process. Meanwhile, the sum of the function $\kappa(v_i, \varsigma)$ with respect to the label v_i can be actually regarded as the matching number constraint and larger value of the sum means more matches.

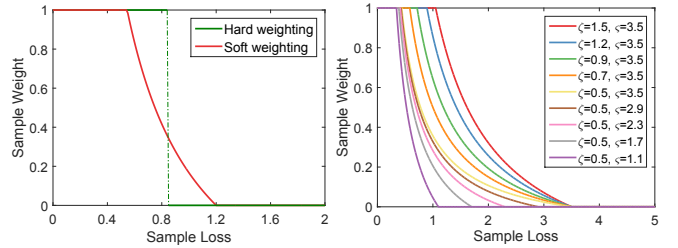


Figure 2: Hard and soft weighting. The left figure is a comparison for soft curve ($\zeta = 1.0, \varsigma = 1.2$) with hard one. The right shows the change curve between sample weight and loss under different parameter ζ, ς .

3.3 Transformation Estimation

The Gaussian kernel is symmetric and positive define. It also makes the regularization term easy to rewrite in form under the Representer theorem. According to the theory of RKMS, the displacement function \mathbf{f} can be expressed as a linear combination of multiple Gaussian kernels in RKHS. More details can be seen in [Ma *et al.*, 2015b]. Then the function \mathbf{f} takes the following specific form of Gaussian radial basis function:

$$\mathbf{f}(\mathbf{x}) = \sum_{i=1}^n \Gamma(\mathbf{x}, \mathbf{x}_i) \mathbf{a}_i, \quad (7)$$

where the coefficient \mathbf{a}_i is a column vector, and the Gaussian kernel $\Gamma(\mathbf{x}_i, \mathbf{x}_j) = \exp\{-\beta\|\mathbf{x}_i - \mathbf{x}_j\|^2\} \cdot \mathbf{I}$. Moreover, the regularization item $\Phi(\mathbf{f})$ is specified as the L_2 norm, i.e.,

$\Phi(\mathbf{f}) = \|\mathbf{f}\|_{\Gamma}^2 = \langle \mathbf{f}, \mathbf{f} \rangle_{\Gamma}$. Then Eq. 3 can be further described as follows:

$$\min_{\mathbf{f}, v} - \sum_{i=1}^n v_i \sum_{r=1}^m w_{ir} \exp \left\{ -\frac{\|\mathbf{x}_i^T - \Gamma_i \mathbf{C} - \mathbf{y}_r^T\|^2}{\sigma^2} \right\} + \sum_{i=1}^n \kappa(v_i, \varsigma) + \lambda \cdot \text{tr}(\mathbf{C}^T \Gamma \mathbf{C}) + \mathfrak{S}, \quad (8)$$

where Γ is the kernel matrix with $\Gamma_{i,j} = \pi(\mathbf{x}_i, \mathbf{x}_j) = \exp \left\{ -\beta \|\mathbf{x}_i - \mathbf{x}_j\|^2 \right\}$, and Γ_i represents the i th row of Γ . $\mathbf{C} = (\mathbf{a}_1, \dots, \mathbf{a}_n)^T$ is the coefficient matrix, and $\text{tr}(\cdot)$ denotes the trace of matrix, and the constant \mathfrak{S} satisfies $\mathfrak{S} = \sum_i v_i \sum_r w_{ir}$.

3.4 Alternating Optimization

Since Eq.8 is not convex, it is almost impossible to find the global minimum by existing algorithms. Whereas Gaussian fields is differentiable and preferably convex in local area around the optimal position, a stable local extremum is enough for most applications. So, the gradient-based numerical approaches (e.g., quasi-Newton method, conjugate gradient method) can be used to solve it by deterministic annealing (DA). As iteration, a stable local minimum can be reached with a better chance [Wang *et al.*, 2017]. There are two variables to optimize in Eq. 8. Here, an alternating optimization is used, namely, optimizing one variable with the other fixed in an alternating way. For v , it is easy to obtain its optimal value. We only need to update it by Eq. 6 during iteration.

To optimize \mathbf{f} , first ignore the items without the \mathbf{f} because the derivatives of these items w.r.t \mathbf{f} are 0. Then the equation becomes the following objective function $E_{\sigma}(\mathbf{C})$:

$$\min_{\mathbf{f}, v} - \sum_{i=1}^n v_i \sum_{r=1}^m w_{ir} \exp \left\{ -\frac{\|\mathbf{x}_i^T - \Gamma_i \mathbf{C} - \mathbf{y}_r^T\|^2}{\sigma^2} \right\} + \lambda \cdot \text{tr}(\mathbf{C}^T \Gamma \mathbf{C}). \quad (9)$$

Since Eq. 9 is always local continuously differentiable w.r.t \mathbf{C} , take its derivative w.r.t parameter \mathbf{C} , then we have

$$\frac{\partial E_{\sigma}(\mathbf{C})}{\partial \mathbf{C}} = \sum_{i=1}^n \sum_{r=1}^m \frac{2w_{ir}}{\sigma} \Gamma_i^T \rho_{ir} \exp \left\{ -\|\rho_{ir}\|^2 \right\} + 2\lambda \Gamma \mathbf{C}, \quad (10)$$

where $\rho_{ir} = \frac{\mathbf{x}_i^T + \Gamma_i \mathbf{C} - \mathbf{y}_r^T}{\sigma}$. Here, Eq. 9 is convex only in the local neighborhood around the optimal solution. Thus, it is almost unlikely to directly get the global minimum. Using the derivative for Eq. 10, we solve the equation by the gradient-based quasi-Newton algorithm with DA.

During the optimization, a rigid-to-nonrigid technique is used to enhance the model convergence by performing DA on scale parameter σ^2 . In other words, by varying σ from a larger initial value, it gradually makes Eq. 9 approach the local area around the real minimum. More specially, we initialize σ with a large value, and gradually update it with $\sigma \mapsto \nu \sigma$ in each iteration, where ν is the annealing rate. Note that the criterion is convex over a large area around the global minimum at a large σ . As σ decreases, the position of the global minimum will tend towards smoothness. Considering that Eq. 9

Algorithm 1 Self-selected point matching

Input: Reference points with descriptors $\{\mathbf{x}_i, F(\mathbf{x}_i)\}_{i=1}^n$, and sensed points with descriptors $\{\mathbf{y}_r, F(\mathbf{y}_r)\}_{r=1}^m$, parameters $\beta, \lambda, \eta, \nu, n_0$, iterations T_{iter}, μ

Output: Optimal transformation τ

- 1: Calculate the Gram matrix Γ and matrix \mathbf{U} ;
- 2: Initialize parameter σ and the coefficient matrix \mathbf{C}_s
- 3: Compute the distance between $\{F(\mathbf{x}_i)\}_{i=1}^n$ and $\{F(\mathbf{y}_i)\}_{i=1}^m$ as the weights $\{w_{ir}\}_{i,r=1}^{n,m}$
- 4: **repeat** (DA)
- 5: $t \leftarrow t + 1$
- 6: Using the quasi-Newton algorithm (e.g. BFGS) to solve Eq. 12 based on $v^{(t-1)}$
- 7: Update the coefficient matrix $\mathbf{C}_s = \arg \min_{\mathbf{C}_s} E_{\sigma}(\mathbf{C}_s)$
- 8: Update v : calculate the $v^{(t)}$ via the Eq. 6
- 9: Using the $\mathbf{C}_s, v^{(t)}$ to calculate the root mean square error $rmse^{(t)}$ of the selected points pairs
- 10: **if** $\varsigma < \varsigma_{max}$ **then**
- 11: $\varsigma \leftarrow \mu \varsigma; \zeta \leftarrow \mu \zeta$
- 12: **end if**
- 13: Update annealing rate $\sigma = \nu \sigma$
- 14: **until** $rmse^{(t)} > \min_{1 \leq \tilde{t} \leq t-1} rmse^{(\tilde{t})}$ or $t \geq T_{iter}$
- 15: $\mathbf{C}_s, v^* = \arg \min_{\tilde{t}} rmse^{(\tilde{t})}$
- 16: Calculate the transformation $\tau(\mathbf{x}) = \mathbf{x} + \mathbf{f}(\mathbf{x})$ by Eq.7
- 17: Return τ ;

is convex in a local small area near its minimum, it is possible to converge to a new global minimum using the global minimum of last iteration as initial value of next iteration. So, the global minimum is more easier to obtain as iteration.

It is complex to directly use all the points in both time and space. A suboptimal but simpler method may be more effective. Kernel approximation is efficient for selecting an RKHS. Low-rank kernel approximation can yield a large speed improvements with little accuracy loss and constraints both the nonrigid transformation and space [Wang *et al.*, 2017]. Specifically, low-rank kernel matrix Γ^s is the closest n_0 -rank matrix Γ , which satisfies the Frobenius norm,

$$\arg \min_{\Gamma^s} \|\Gamma - \Gamma^s\|_F, \quad s.t. \text{rank}(\Gamma^s) \leq n_0. \quad (11)$$

Using the eigenvalue decomposition of Γ , the approximated kernel matrix can be taken as $\Gamma^s = \mathbf{P} \mathbf{\Lambda} \mathbf{P}$, where $\mathbf{\Lambda}$ is an $n_0 \times n_0$ diagonal matrix with n_0 largest eigenvalue and \mathbf{P} is an $n \times n_0$ matrix with the eigenvectors. Finally, the regularized Gaussian criterion (9) and its derivative changes as

$$\min_{\mathbf{f}, v} - \sum_{i=1}^n v_i \sum_{r=1}^m w_{ir} \exp \left\{ -\frac{\|\mathbf{x}_i^T - \mathbf{U}_i \mathbf{C}_s - \mathbf{y}_r^T\|^2}{\sigma^2} \right\} + \lambda \cdot \text{tr}(\mathbf{C}_s^T \mathbf{U} \mathbf{C}_s), \quad (12)$$

$$\frac{\partial E_{\sigma}(\mathbf{C}_s)}{\partial \mathbf{C}_s} = \sum_{i=1}^n \sum_{r=1}^m \frac{2w_{ir}}{\sigma} \mathbf{U}_i^T \tilde{\rho}_{ir} \exp \left\{ -\|\tilde{\rho}_{ir}\|^2 \right\} + 2\lambda \mathbf{U} \mathbf{C}_s, \quad (13)$$

where $\tilde{\rho}_{ir} = \frac{\mathbf{x}_i^T + \Gamma_i \mathbf{C}_s - \mathbf{y}_r^T}{\sigma}$, \mathbf{C}_s denotes the new $n_0 \times d$ coefficient matrix with elements $(a_1, \dots, a_{n_0})^T$, $\mathbf{U}_{n \times n_0} = \mathbf{P}\mathbf{A}$ and \mathbf{U}_i represents the i th row of \mathbf{U} . We summarise the optimization procedure in Algorithm 1.

Note that the time complexity of Gaussian fields criterion is about $\mathcal{O}(n^2m + n^3)$. However, the time complexity reduces to $\mathcal{O}(nm)$ by the low-rank kernel matrix approximation.

4 Experiments

In this section, we assess our approach by comparing with several methods on different datasets, i.e., the synthesized 2D shapes dataset [Zheng and Doermann, 2006], the benchmark IMM Face Databaset¹ and the Oxford dataset². Since different methods are proposed for different applications, we select the targeted methods on different datasets for a fair comparison. For the 2D shape and 3D face datasets, we make comparisons with the typical methods, mainly including CPD [Myronenko and Song, 2010], L2E [Ma *et al.*, 2015b], RGF [Ma *et al.*, 2015c], PR-GLS [Ma *et al.*, 2016] and context-aware gaussian fields (CA-GF) [Wang *et al.*, 2017]. On the Oxford dataset, some latest algorithms based on RANSAC [Lowe, 2004] are made comparisons, such as FSC [Wu *et al.*, 2015] and the imprecise points removing algorithm (IPRA) [Wu *et al.*, 2015]. All methods are implemented in Matlab, and tested on the same environment.

Parameter Setting In the experiments, a deterministic annealing technique strategy is applied on the scale parameter σ^2 to improve the algorithm convergence. More specially, give a large initial value of σ^2 , and reduce them with a fixed annealing rate ν by $\sigma = \nu\sigma$. We empirically set $\sigma = 2$, and $\nu = 0.93$ throughout the whole experiments. The parameter β identifies the range width of the interaction between points. Since the points have been normalized to 0 mean and unit variance, β would be similar to different samples, and we set $\beta = 0.2$ in the experiments. The parameter of regularization term includes λ which works for trade-off the smoothness, and λ is affected by the degree of data degradation, which is set by $\lambda = 0.1$. Note that parameter η reflects the effect of feature descriptor on the model. So, it can be set as 1 for the same kind of data. Besides, we initialize \mathbf{C}_s as zero vector.

Whereas the matching error is large at the beginning of iteration, the tolerance of the model to the error would be large for selecting more points. So, we dynamically update ζ , ς that decide the tolerance of the model to the error. As analyzed in Fig.1, we give a large initial value of ζ , ς and reduce them with a fixed rate μ . Since the optimization would be terminated, we choose a lower bound of ζ and set $\zeta_{final} = 1.0$, and also set a lower bound $\varsigma_{final} = 1.2$ to control the weight of confused points. ς is fixed when $\varsigma \leq \varsigma_{final}$.

Low-rank kernel matrix approximation is applied to reduce the computational cost. Parameter n_0 is the number of the selected eigenvalues, and as a trade-off, it controls the balance between runtime and matching accuracy. The proposed method is best for $n_0 \in [15, 20]$, where eigenvector \mathbf{P} and eigenvalue are calculated by the fast Gauss transform (FGT)

and its experimental analysis is shown in Fig.3. From it, we set $n_0 = 15$ in the experiments.

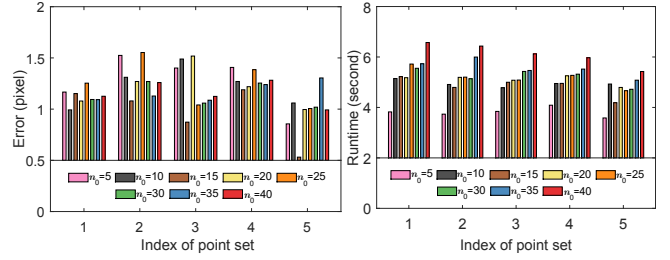


Figure 3: Matching results on IMM face landmarks for different n_0 . Left figure shows matching errors and the right is the runtime.

4.1 Results on Non-rigid Synthesized Data

Fig.4 gives the matching results of our method on point sets with deformation, noise, outliers and rotation. In each group, four degradation levels are designed to test the robustness of our method. The intuitive experimental results in Fig.4 show that the model point sets are well aligned onto the data sets except for the data sets distorted by noises.

Moreover, for the data with different deformations, noises, outliers and occlusion, we list the visual results of the typical methods, e.g., CPD, L2E, RGF, PR-GLS and the latest CA-GF (see Fig.5). From it, the shapes badly eroded by noises or outliers, are still well-aligned by our method. Especially for the occlusion, other methods almost fail or only align part of the shape while our method nearly gets a complete alignment. It owns to our soft selection. Initially, the points with large error are blocked by the soft labels, which makes the model sequential to select points, not to disturb other points. More reliable points are first selected to gradually enhance the model until it converges. Even if later points are noises or outliers, they are not enough to affect the estimated model. As presented in Table.1, our method has lower errors.

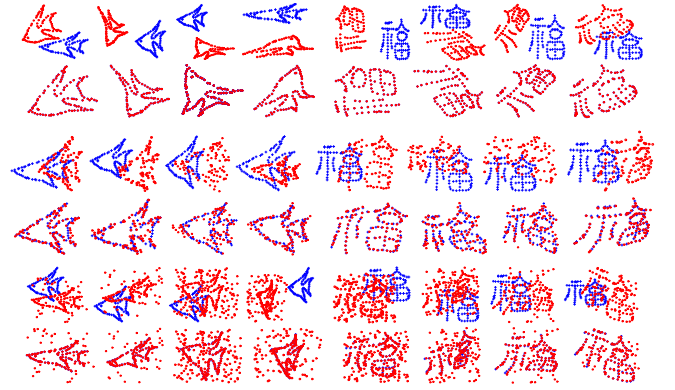


Figure 4: The results of our method on the synthesized data with different degenerations in every two rows. For each group, the upper is the input points while the lower is the matching results.

4.2 Registration on IMM Face Dataset

In this section, we test our method on the IMM Face dataset. It contains 240 images with 640×480 resolution on 40 human faces. For this data, each face has 6 samples and each

¹Available at: <http://www.imm.dtu.dk/aam/datasets/datasets.html>

²Available at: <http://www.robots.ox.ac.uk/vgg/data/dataaff.html>

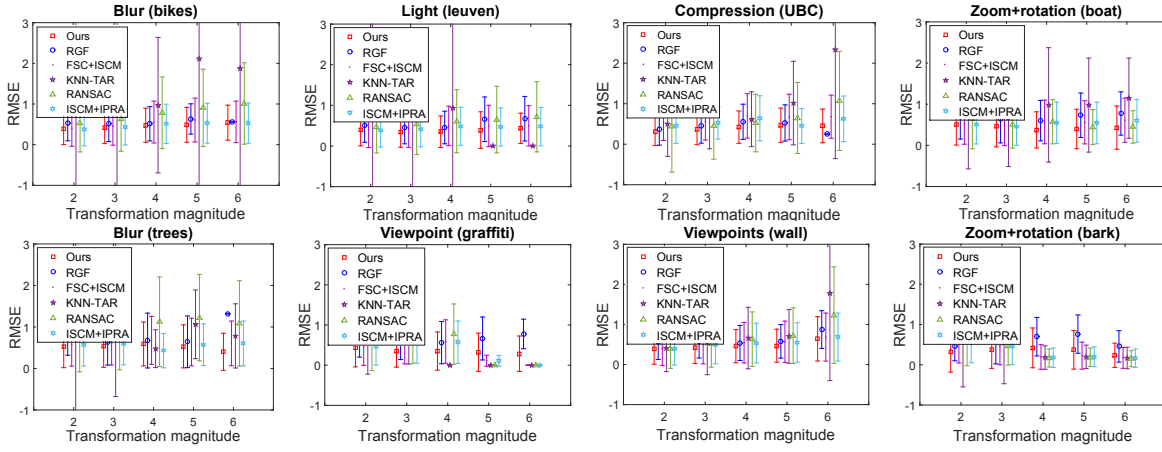


Figure 6: RMSE values of the images with different transformation magnitudes using different methods on the Oxford dataset. The first image of each group in the Oxford dataset is taken as the reference image and others correspond the magnitude 2-6, respectively.

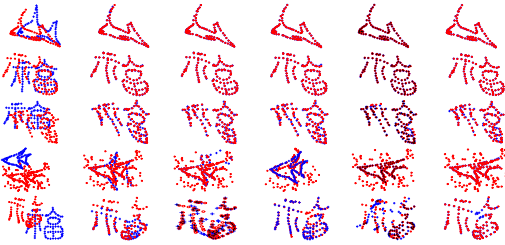


Figure 5: Schematic results of different methods. The first column shows the original data, marked as Def1, Def2, Noi, Out, Occ from top to bottom. The next 2-6 columns list the results of CPD, L2E, PR-GLS, CA-GF and ours from left to right.

Method	Def1	Out	Def2	Noi	Occ
CPD	0.339	0.420	0.076	0.150	0.291
RGF	0.190	0.094	0.070	0.073	0.048
L2E	0.090	0.117	0.003	0.044	0.050
PR-GLS	0.017	0.025	0.004	0.018	0.022
CA-GF	0.024	0.007	0.017	0.111	0.06
Ours	0.008	0.020	0.003	0.014	0.006

Table 1: Comparisons of matching errors for CPD, RGF, L2E, PR-GLS, CA-GF, and our method on above data.

sample contains 58 landmarks. In the experiment, our goal is to align the model points to the data landmarks. Due to complex facial expression, the deformation is nonrigid and also serious in local area. Aligning such points is difficult. However, Gaussian kernel can approach the real deformation better as shown in Fig.7. From the details of Fig.7, RGF can align the edges well, but it must be correct to single point for PSR. Since some landmark pairs are not the actual correspondences in real face, it is almost unlikely to align all the points. Such points indeed affect CPD, RGF and PR-GLS, and only obtains a rough alignment. As expected, our method selects more reliable points, and achieves an accurate alignment.

4.3 Registration on Real Image

The Oxford dataset is a widely used dataset in image registration. It consists of more than 40 image pairs with different

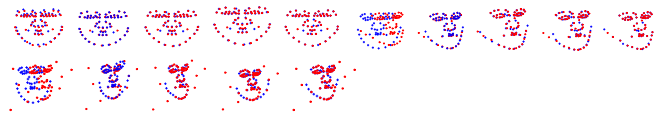


Figure 7: Comparisons for CPD, RGF, PR-GLS and our method on the IMM Face Dataset (from left to right).

variations. Fig.6 shows the results on the Oxford dataset. To test the validity of our method in real scenario, we further employ our approach to match the real images with complex deformations, which are captured at different time from different viewpoints and contain large scale changes, local self-similar texture even nonrigid distortions (see Fig.8). In this experiment, RGF, K nearest neighbor with the triangle-area representation (KNN-TAR) [Zhang *et al.*, 2014], RANSAC, FSC and IPRA, is used to make comparisons.

5 Conclusion

In this paper, a robust nonrigid Gaussian fields criterion with self-selection is proposed for point matching. The main idea of the proposed method is a novel non-rigid transformation estimation with adaptive self-selected mechanism, and the biggest difference with traditional density estimation methods is that our method alters the convergence manner of model by adaptive sample selection. Moreover, an equivalent matching number constraint is embedded into the nonrigid criterion, which enhances the reliability of point matching. Experiments on three matching datasets show that our method has higher registration accuracy and correct matches than state-of-the-art approaches.

Acknowledgments

This work was supported by the National Key R&D Program of China under Grant 2017YFB1002202, National Natural Science Foundation of China under Grant 61773316 and 61379094, and the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences.

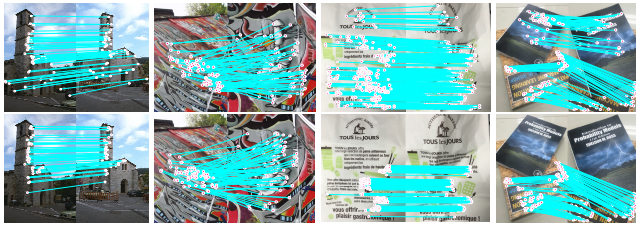


Figure 8: Matching examples on real images. The top row is the results of our method and the results of IPRA are shown below.

References

- [Bai *et al.*, 2018] Lifei Bai, Xianqiang Yang, and Huijun Gao. Nonrigid point set registration by preserving local connectivity. *IEEE transactions on cybernetics*, 48(3):826–835, 2018.
- [Chou and Wang, 2015] Chih Chung Chou and Chieh-Chih Wang. 2-point ransac for scene image matching under large viewpoint changes. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 3646–3651. IEEE, 2015.
- [Hong and Kim, 2016] Seonghun Hong and Jinwhan Kim. Efficient visual slam using selective image registration for autonomous inspection of underwater structures. In *Autonomous Underwater Vehicles (AUV), 2016 IEEE/OES*, pages 189–194. IEEE, 2016.
- [Jian and Vemuri, 2011] Bing Jian and Baba C Vemuri. Robust point set registration using gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1633–1645, 2011.
- [Jiang *et al.*, 2014] Lu Jiang, Deyu Meng, Teruko Mitamura, and Alexander G Hauptmann. Easy samples first: Self-paced reranking for zero-example multimedia search. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 547–556. ACM, 2014.
- [Kahaki *et al.*, 2016] Seyed Mousavi Kahaki, Md Jan Nordin, Amir H Ashtari, and Sophia J Zahra. Deformation invariant image matching based on dissimilarity of spatial features. *Neurocomputing*, 175:1009–1018, 2016.
- [Litman *et al.*, 2015] Roei Litman, Simon Korman, Alexander Bronstein, and Shai Avidan. Inverting ransac: Global model detection via inlier rate estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5243–5251, 2015.
- [Lowe, 2004] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [Ma *et al.*, 2015a] Jiayi Ma, Yong Ma, Ji Zhao, and Jinwen Tian. Image feature matching via progressive vector field consensus. *IEEE Signal Processing Letters*, 22(6):767–771, 2015.
- [Ma *et al.*, 2015b] Jiayi Ma, Weichao Qiu, Ji Zhao, Yong Ma, Alan L Yuille, and Zhuowen Tu. Robust l2e estimation of transformation for non-rigid registration. *IEEE Trans. Signal Processing*, 63(5):1115–1129, 2015.
- [Ma *et al.*, 2015c] Jiayi Ma, Ji Zhao, Yong Ma, and Jinwen Tian. Non-rigid visible and infrared face registration via regularized gaussian fields criterion. *Pattern Recognition*, 48(3):772–784, 2015.
- [Ma *et al.*, 2016] Jiayi Ma, Ji Zhao, and Alan L Yuille. Non-rigid point set registration by preserving global and local structures. *IEEE Transactions on Image Processing*, 25(1):53–64, 2016.
- [Myronenko and Song, 2010] Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275, 2010.
- [Pinheiro *et al.*, 2017] Miguel Amável Pinheiro, Jan Kybic, and Pascal Fua. Geometric graph matching using monte carlo tree search. *IEEE transactions on pattern analysis and machine intelligence*, 39(11):2171–2185, 2017.
- [Qu *et al.*, 2017] Han Bing Qu, Jia Qiang Wang, Bin Li, and Ming Yu. Probabilistic model for robust affine and non-rigid point set matching. *IEEE transactions on pattern analysis and machine intelligence*, 39(2):371–384, 2017.
- [Tsin and Kanade, 2004] Yanghai Tsin and Takeo Kanade. A correlation-based approach to robust point set registration. In *European conference on computer vision*, pages 558–569. Springer, 2004.
- [Wang *et al.*, 2017] Gang Wang, Qiangqiang Zhou, and Yufei Chen. Robust non-rigid point set registration using spatially constrained gaussian fields. *IEEE Transactions on Image Processing*, 26(4):1759–1769, 2017.
- [Wu *et al.*, 2015] Yue Wu, Wenping Ma, Maoguo Gong, Linzhi Su, and Licheng Jiao. A novel point-matching algorithm based on fast sample consensus for image registration. *IEEE Geoscience and Remote Sensing Letters*, 12(1):43–47, 2015.
- [Zhang *et al.*, 2014] Kai Zhang, XuZhi Li, and JiuXing Zhang. A robust point-matching algorithm for remote sensing image registration. *IEEE Geoscience and Remote Sensing Letters*, 11(2):469–473, 2014.
- [Zheng and Doermann, 2006] Yefeng Zheng and David Doermann. Robust point matching for nonrigid shapes by preserving local neighborhood structures. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):643–649, 2006.
- [Zheng *et al.*, 2017] Qingqing Zheng, Yi Wang, and Pheng Ann Heng. Online robust image alignment via subspace learning from gradient orientations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1753–1762, 2017.
- [Zhou *et al.*, 2015] Xiaowei Zhou, Menglong Zhu, and Kostas Daniilidis. Multi-image matching via fast alternating minimization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4032–4040, 2015.