Artificial Intelligence Review
https://doi.org/10.1007/s10462-022-10292-4

# Non-rigid point set registration: recent trends and challenges

**Xiaohui Yuan**[1] · **Amar Maharjan**[1]



**Abstract**
Non-rigid point set registration has been used in a wide range of computer vision applications such as human movement tracking, medical image analysis, three dimensional (3D) object reconstruction and is a very challenging task. It has two fundamental tasks. One is to find correspondences between two or more point sets and another is to transform a point set so that it aligns with other point sets. There has been significant progress in the past two decades in the non-rigid registration field but it still has major challenges and is an active research area in the computer vision and pattern recognition community. In this review, we present a survey of non-rigid point set registration. Unlike recent surveys, we focus on the mathematical foundations of non-rigid registration methods, categorize the methods from several perspectives, and discuss open challenges. We categorize the methods according to correspondence models, motivations, and challenges such as deformation, data degradation, computational efficiency, and different constraints used in the methods to achieve accurate registration results. We present the publicly available data sets and different evaluation techniques employed in the methods. Further, we discuss open challenges, recent trends, and potential directions for future work in this area.

**Keywords** Registration · Point Set · Point Cloud · Optimization · Deformation

## 1 Introduction

Non-rigid point set registration finds correspondences between points to derive a transformation function that aligns the point sets. Sources of such point sets could be directly acquired by imaging devices or post-processed from images and videos. For example, three-dimensional (3D) point sets representing the objects in indoor applications can be captured with Kinect and Intel realsense devices. In addition, Lidar devices, widely used in land surveys and autonomous vehicles, capture outdoor environments and generate large 3D point clouds. The increasing capability and application of acquiring point sets foster the development of point set registration methods.

In point set registration, there could be two or more point sets involved. One point set is transformed so that it aligns with the other point set as close as possible. The transformed

✉ Xiaohui Yuan
xiaohui.yuan@unt.edu

[1] University of North Texas, Denton, TX, USA

Published online: 11 October 2022

point set is referred to as *model* or *template* point set whereas the fixed point set is often called *scene* or *target* point set. If the registration contains more than two point sets, usually called groupwise registration, it is often solved by pairwise registration with either one-versus-all or sequential strategy. Point set registration is a fundamental task in many computer vision and pattern recognition applications such as transfer of knowledge (Kong et al. 2018; Lu et al. 2019b; Yuan et al. 2018), pose estimation (Ye et al. 2016; Yuan et al. 2017), medical image analysis (Kolesov et al. 2016), surface reconstruction (Lu et al. 2019a), three-dimensional object reconstruction (Wang et al. 2016b), and human movement tracking (Jian and Vemuri 2011; Maharjan and Yuan 2022; Qu et al. 2017).

Non-rigid point set registration is an ill-posed problem. For each point, there could be a number of transformations, which makes it a very challenging task. In practice, constraints are employed to achieve a unique solution. Examples of such constraints include moving local neighborhood points in a similar direction and ensuring the smoothness of the transformation function. A different number of points between the point sets, significant deformation between the point sets, degradation in the point sets such as noises, outliers, occlusions, and large computational costs due to a huge number of points in the point sets make the registration tasks even more difficult.

There are review articles that cover some point set registration methods. Tam et al. (2013) covered both rigid and non-rigid registration methods from the perspective of data fitting. The review focused on 3D point clouds and meshes. Maiseli et al. (2017) discussed point set registration developments and trends. The paper summarizes and reviews methods up to 2017. Zho et al. (2019a) focused on the registration methods from the viewpoint of pairwise registration and groupwise registration.

Despite the aforementioned efforts and in the light of the fast growth of the field, this paper summarizes recent research results of non-rigid point set registration methods and presents a perspective with a coherent evaluation of the trends. Different from the existing surveys on non-rigid registration methods, this study discusses the existing methods from

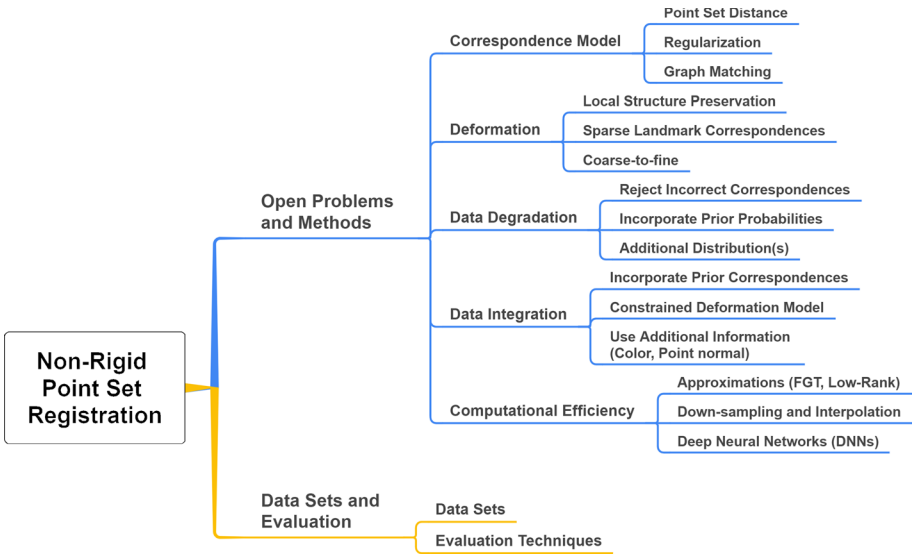


**Fig. 1** An overview of the organization of this survey

two aspects: open problems and evaluation of non-rigid point set registration. Figure 1 shows the organization of this survey. We aim at providing researchers with a perspective view on the challenges, techniques, and evaluation. Although point sets sometimes refer to manually or automatically extracted salient points from images, our focus is on the problems and techniques that deal with range data that provide 3D information of the object of interest and are large in volume, e.g., data from LiDAR, Time-of-Flight, and Structured Light sensors. The processing of unstructured 3D point collection often faces unique problems that are non-existent in the extracted salient points such as uneven deformation. In addition, problems that fall into the scope of rigid transformation, e.g., registering 3D scans of building structures and objects, are not included in our discussion.

The rest of this paper is organized as follows: Sect. 2.1 presents different types of correspondence models and regularization terms used in the existing methods in detail. Section 2.2 discusses the challenges of handling large deformation, and of different strategies used by the methods to deal with it. Section 2.3 discusses data degradation such as noises, outliers, occlusions in the point sets and categorizes the methods based on how they handle the data degradation to reduce its effects. Section 2.4 reviews different approaches to incorporate additional knowledge and constraints in the methods. Computational efficiency in the existing methods is discussed in Sect. 2.5. In Sect. 3, we present publicly available data sets and discuss evaluation techniques used in the methods. Finally, Sect. 4 concludes with a discussion of the existing challenges, trends, and future work.

## 2 Open problems and methods

### 2.1 Correspondence model

A key component of a registration method is the correspondence model. In general, non-rigid point set registration minimizes the overall distance between point sets with an objective function including two components: a distance function and a regularization term:

$$E(\mathbf{X}, \mathbf{Y};\theta) = D(\mathbf{X}, \mathbf{Y};\theta) + \lambda R(\theta) \tag{1}$$

where $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ and $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M\}$ denote point sets, $\theta$ denotes model parameters, and $\lambda$ is a weight for the regularization term $R(\theta)$. The regularization term ensures that plausible correspondences are achieved instead of reaching an erroneous minimum distance.

#### 2.1.1 Point set distance

The distance function usually takes the form of

$$D(\mathbf{X}, \mathbf{Y};\theta) = \sum_{i,j} w_{ij} d(\mathbf{x}_i, f(\mathbf{y}_j;\theta)), \tag{2}$$

where $w_{ij}$ is a weight for points $\mathbf{x}_i$ and $\mathbf{y}_j$ and $\sum_{i,j} w_{ij} = 1$, $d(\mathbf{x}_i, \mathbf{y}_j)$ returns the distance between point $\mathbf{x}_i$ and the transformed point $\mathbf{y}_j$ using function $f$. An intuitive distance function is Euclidean distance that takes the form of $\left\|\mathbf{x}_i - \mathbf{y}_j\right\|_2^2$ (Chui and Rangarajan 2000b, 2003; Lian et al. 2017). To compute the distance, it is necessary to assign correspondence among points. In non-rigid point set registration, a point can be mapped to multiple points

with a probability, i.e., soft assignment (Hinton et al. 1992; Revow et al. 1996). In such methods, registration is usually formulated as density estimation based on Gaussian mixture model (GMM) (Myronenko and Song 2010; Myronenko et al. 2007). Chui and Rangarajan (2000a) proposed a mixture point matching (MPM) method that formulates the point set registration as a mixture density estimation problem. The mean of each Gaussian model is specified by a point and a fixed covariance is used. The models are transformed to maximize the posterior probability such that the GMM centroids are aligned with the points in the other set. Following this idea, Zeng et al. (2017b) used the mixture-feature Gaussian mixture model (MGMM) for correspondence and employed global-local spatial constraints during the transformation.

Besides GMM, Student's T model was used (Zhou et al. 2017, 2018, 2014). The Student's T models have heavier tails compared to Gaussian models, which tend to produce values that fall far from the mean and, hence, help mitigate data degradation such as noise and outliers. Fu and Zhou (2016) used a mixture of asymmetric Gaussian mixture model (MoAG) for density estimation. Unlike GMM, each Gaussian model in the MoAG is asymmetric, which provides a more suitable model for the spatially asymmetric distribution. The MoAG model demonstrated improved robustness to data degradation such as noises, outliers, and occlusions (Fu and Zhou 2016).

Chui and Rangarajan (2000b, 2003) proposed a general framework method based on the robust point matching (RPM) algorithm (Rangarajan et al. 1997), which uses weight in the range of [0, 1] to indicate the correspondence. The methods employ thin plate spline (TPS) to regulate the transformation function to achieve smoothness in transformation. Many registration methods extended the RPM method to deal with outliers (Yang 2011) or global optimal solution (Lian et al. 2017).

Another class of methods that align point sets without explicitly establishing correspondences is presented in Bing and Vemuri (2005); Glaunes et al. (2004); Jian and Vemuri (2011); Tsin and Kanade (2004). In these methods, each point set is modeled by a probability distribution, and the statistical discrepancy between the two distributions is minimized by finding the optimal transformation function. Wang et al. (2008) used Jensen–Shannon (JS) divergence to simultaneously register multiple point sets, which are represented by the Gaussian mixture models. The JS divergence is used to quantify the difference (or similarity) between multiple probability distributions. Given the probability distributions $\{p_i\}_{i=1}^{n}$, the JS divergence of $p_i$ is given by,

$$JS(p_1, ..., p_n) = H\Big(\sum \pi_i p_i\Big) - \sum \pi_i H(p_i) \tag{3}$$

where $\pi_i = \{\pi_1, ..., \pi_n | \pi_i > 0, \sum \pi_i = 1\}$ are the weights of the distributions $p_i$, and $H(\cdot)$ computes the entropy. Similarly, a registration framework is proposed where the point set registration is treated as minimizing the discrepancy between the Gaussian mixture models (Bing and Vemuri 2005; Jian and Vemuri 2011). In this framework, $L2$ distance measures the similarity between the Gaussian mixture densities, $P_X$ and $P_Y$:

$$D(P_X(\mathbf{x}), P_Y(\mathbf{x};\theta)) = \int (P_X(\mathbf{x}) - P_Y(\mathbf{x};\theta))^2 \, dx \tag{4}$$

This framework is generic and can interpret several point set registration methods meaningfully such as TPS-RPM (Chui and Rangarajan 2000b), MPM (Chui and Rangarajan 2000a), kernel correlation-based registration method (Tsin and Kanade 2004), GMM (Bing and Vemuri 2005), and CPD (Myronenko and Song 2010; Myronenko et al. 2007).

Instead of using overall global distance metrics between the point sets, some methods maintain primarily the local neighborhood structures in the cost functions (Ma et al. 2017, 2019b; Zheng and Doermann 2006). Zheng and Doermann (2006) introduced the mutual distance function

$$D(\mathbf{X},\mathbf{Y},\theta)=\sum_{i=1}^{N}\sum_{\mathbf{x}_n\in N_i} d(f(\mathbf{x}_i),f(\mathbf{x}_n))+\sum_{j=1}^{M}\sum_{\mathbf{y}_m\in N_j} d(f^{-1}(\mathbf{y}_j),f^{-1}(\mathbf{y}_m)) \tag{5}$$

where $N_k$ is a set which contains neighborhood points of point $k$, $d(i,j)=0$ if $j\in N_i$ otherwise 1, $f(\cdot)$ is match (correspondence) function and $f^{-1}(\cdot)$ is inverse match function. The first summation computes the total distance of the mapped points in $\mathbf{Y}$ for each point to its neighbors in $\mathbf{X}$. The second summation computes the total distance of the mapped points in $\mathbf{X}$ for each point to its neighbors in $\mathbf{Y}$. The distance function $d(i,j)=0$ if $j\in N_i$; otherwise $d(i,j)=1$. This cost function aims to achieve a stable local relationship among the neighboring points and preserve the neighborhood distances under deformation. To minimize the cost function, neighboring points of a point matches the neighboring points of the matched point. That is, the optimal match is obtained when the local neighborhood structure is intact even if the point set is deformed. Similar to this idea, Ma et al. (2017, 2019b) presented a locality preserving matching (LPM) method to remove mismatches from putative correspondences. The true correspondences should maintain consensuses on both neighboring points and neighborhood topology; whereas the false correspondences typically fail to maintain the consensus and will be rejected. The consensus of the neighborhood topology is based on the ratio of length and the angle between vectors $\mathbf{v_i}$ and $\mathbf{v_j}$, which are constructed from the corresponding points $(\mathbf{x_i},\mathbf{y_i})$ and $(\mathbf{x_j},\mathbf{y_j})$, respectively. Table 1 summarizes the commonly used distance functions in registration methods.

**Table 1** Commonly used distance functions in registration methods

| Distance | Method |
|---|---|
| $\sum_{i,j} w_{ij} d(\mathbf{x}_i,f(\mathbf{y}_j,\theta))$ | Chui and Rangarajan (TPS-RPM) (2000b, 2003), Chui and Rangarajan (MPM) (2000a), Myronenko and Song (CPD) (2010); Myronenko et al. (2007), Ge and Fan (GLTP) (2019); Ge et al. (2014), Tsin and Kanade (KC) (2004), Yang (2011), Fu and Zhou (MoAG) (2016), Zhang et al. (MGMM) (2017b), Zhou et al. (2014, 2017, 2018), Lian et al. (APM) (2017), Yang et al. (2018), Calvo et al. (CCPD) (2018), Zhang et al. (SCCD) (2017a), Ge et al. Song and Fan (2015), Maharjan and Yuan (2020); Maharjan et al. (2021), Zhu et al. (2019b), Ge et al. (LSP) Ge and Fan (2015), Wang et al. (SCGF) Wang et al. (2017), Panaganti et al. Panaganti and Aravind (2015) Zhou et al. (DSMM) (2017), Zhou et al. (2014) |
| $\sum_i w_i d(\mathbf{x}_i,f(\mathbf{y}_i,\theta))$ | Ma et al. (VFC) (2014), Ma et al. (MS-RPM) (2019a), Yang et al. (2015), Wang et al. (MSCVF) (2016a, 2018) |
| $\sum_{i=1}^{N}\sum_{\mathbf{x}_n\in N_i} d(f(\mathbf{x}_i),f(\mathbf{x}_n))$ $+\sum_{j=1}^{M}\sum_{\mathbf{y}_m\in N_j} d(f^{-1}(\mathbf{y}_j),f^{-1}(\mathbf{y}_m))$ | Zheng and Doermann (2006), Ma et al. (LPM) (2017, 2019b) |
| $\int (P_X(\mathbf{x})-P_Y(\mathbf{x},\theta))^2\, d\mathbf{x}$ | Tsin and Kanade (KC) (2004), Bing and Vemuri (GMMReg) (2005); Jian and Vemuri (2011) |
| $H(\sum \pi_i p_i)-\sum \pi_i H(p_i)$ | Wang et al. (2008) |

### 2.1.2 Regularization

Typical regularization strategies include smoothness of the transformation, motion coherence, and preservation of local neighborhood structure. To maintain the topological structure of the point set during the transformation, Myronenko et al. (2007) proposed Coherent Point Drift (CPD) based on the motion coherence theory (Yuille and Grzywacz 1988). The basic idea is that points that are close to each other should move coherently in the same/similar direction. The regularization term is the norm of the transformation function in the Reproducing Kernel Hilbert Space (RKHS) (Chen and Haykin 2002) as follows:

$$\|f\|_{\mathbb{H}_K}^2 = \int_{\mathbb{R}^D} \frac{|\hat{f}(\mathbf{s})|^2}{\hat{G}(\mathbf{s})}\, d\mathbf{s} \tag{6}$$

where $G$ is a kernel function and $\hat{G}$ is its Fourier transformation. $\hat{f}$ is the Fourier transformation of $f$, and $\mathbf{s}$ is the frequency variable. The regularization term measures the oscillatory behavior of a function for smoothness. That is, one function is smoother than another if it oscillates less, which results in a coherent transformation within a spatially adjacent group of points. In the frequency domain, a function is smoother if it has less energy at high-frequency components. In CPD, $G$ takes the form of a Gaussian function.

Hirose (2021) presented a Bayesian formulation of the CPD that uses a prior distribution of displacement vectors for motion coherence. Zho et al. (2019b) extended the CPD by integrating local geometry constraints including local neighborhood and Laplacian coordinate. The local neighborhood constraint ensured the preservation of the local neighborhood after transformation (the key idea is taken from the local linear embedding (LLE), i.e., representing each point by the linear combination of its neighboring points). The Laplacian coordinate was used to preserve the scale of neighborhood structure [see Eqs. (10) and (11)].

A similar idea is used in manifold regularization-based methods (Ma et al. 2019a; Wang et al. 2016a). Manifold regularization is a technique to constrain the function by exploiting the geometry of the probability distribution (Belkin et al. 2006). Let $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{l}$ denote a set of labeled points (i.e., corresponding point pairs) and $\{(\mathbf{x}_j)\}_{j=l+1}^{j=l+u}$ denote a set of unlabeled points. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are close, the conditional probabilities $p(\mathbf{y}|\mathbf{x}_1)$ and $p(\mathbf{y}|\mathbf{x}_2)$ are also similar. Manifold regularization is expressed as follows:

$$\int_{\mathbf{x}\in\mathcal{M}} \left\|\nabla_{\mathcal{M}} f\right\|^2 p(\mathbf{x})\, d\mathbf{x}, \tag{7}$$

where $\nabla_{\mathcal{M}}$ is the gradient of $f$ along the manifold $\mathcal{M}$, measure the smoothness of function $f$. When points are dense, the transformation function should be smoother. That is, the gradient $\nabla_{\mathcal{M}}$ should be small on large marginal probability density $p(\mathbf{x})$. In practice, manifold regularization is approximated with the weighted sum of the squared difference of the transformed points as follows (Belkin and Niyogi 2008; Hein et al. 2005):

$$\frac{1}{(u+l)^2} \sum_{i,j=1}^{l+u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2, \tag{8}$$

where $w_{ij}$ denotes edge weights in the data adjacency graph.

The manifold regularization has been used to deal with data degradation and outlier (Ma et al. 2019a; Panaganti and Aravind 2015; Wang et al. 2016a, 2017). Panaganti and Aravind

(2015) used graph Laplacian regularization to perform registration in the presence of data degradations.Wang et al. (2016a) proposed a manifold regularized coherent vector field (MRCVF) method. The initial correspondences (labeled points) were generated using SIFT keypoints and graph Laplacian regularization was used to learn coherent vector fields. Wang et al. (2017) presented a spatially constrained Gaussian fields (SCGF) method to estimate the transformation function using putative correspondences. The putative correspondences were obtained by matching Shape Context, which was used to estimate the transformation function using manifold regularization. Ma et al. (2019a) proposed a robust point matching method using manifold regularization (MR-RPM), to learn the transformation function. The point set registration is cast as a semi-supervised learning problem where the initial correspondences are computed by matching local feature descriptors such as FPFH (Rusu et al. 2009). The objective function exploits the intrinsic geometrical structures of the point set to derive the transformation function.

The local neighborhood structure is often represented as the weighted sum of the neighbors, which has been used as a constraint to the objective function (Ge and Fan 2015, 2019; Ge et al. 2014). The idea is to retain the weight after the transformation. That is, the distance of all points and their weighted neighborhood structure is minimized:

$$\sum_{j=1}^{M} \left\| f(\mathbf{y}_j;\theta) - \sum_{\mathbf{y}_k \in \tilde{\mathbf{Y}}_j} w_{jk} f(\mathbf{y}_k;\theta) \right\|_2^2, \tag{9}$$

where $\tilde{Y}_j$ is a set that contains K-nearest neighbors of point $y_j$ and $w_{jk}$ is the weight for the neighboring point $y_k$.

Local Structure Preservation (LSP) (Song and Fan 2015) extends the above constraint by including Laplacian Coordinate (LC) to preserve local neighborhood structure and scale. The point set is represented with a graph (V, E), where V and E denote the set of vertices (points) the edges, respectively. The LC constraint minimizes the scale difference during transformation as follows:

$$\mathcal{C}(v_i) = \sum_{(v_i,v_j)\in E} a_{ij}(v_i - v_j), \tag{10}$$

where $a_{ij} \in A$, which is the adjacency matrix to express if two vertices are direct neighbors using zeros and ones, and $(v_i, v_j)$ is an edge between vertices $v_i$ and $v_j$. The regularization term is as follows:

$$\sum_{m=1}^{M} \left\| \mathcal{C}(y_m) - \mathcal{C}(f(y_m;\theta)) \right\|_2^2. \tag{11}$$

Minimizing LC constraint ensures the scale of the neighborhood of points remains the same after transformation.

Zeng et al. (2017b) employed both motion coherence and local structure descriptor as regularization to the objective function. The local structure constraint (LSC) aims to minimize the discrepancy between putative corresponding points, denoted with $\hat{x}$, and the transformed points $f(Y, \theta)$:

$$\sum_{m-1}^{M} \left\| F(\hat{\mathbf{x}}_m) - F(f(\mathbf{y}_m;\theta)) \right\|_2^2 \tag{12}$$

where $F(\cdot)$ computes the summation of local vectors among the nearest neighbors. Let $\mathbf{x}_i$, $i = \{1, \ldots, K\}$, be the nearest neighbors of point $\mathbf{x}_0$. The local vectors originated at $\mathbf{x}_0$ are represented as $\mathbf{v}_i = \overrightarrow{\mathbf{x}_0\mathbf{x}_i}$. The local structure descriptor is computed as follows

$$F(\mathbf{x}_0;K) = \sum_{i=1}^{K} c_i \mathbf{v}_i,$$

where $c_i = \exp\left(-\|\mathbf{v}_i\|^2/\sigma_i^2\right)$ and $\sigma_i^2$ is the covariance. Figure 2 illustrates the computation of the local structure descriptor of a point (depicted as a solid red circle).

Another technique is to employ constraints in the transformation model instead of using regularization so that the point sets follow certain physical properties. Examples of such physical constraints include avoiding mapping two or more points to the same point and alignment of landmarks (Kolesov et al. 2016, 2013a, 2013b; Maharjan and Yuan 2020; Maharjan et al. 2021).

To deal with large and uneven deformation, Maharjan and Yuan (2020); Maharjan et al. (2021) implemented stochastic neighbor embedding (SNE) and landmarks correspondences. The SNE constraint penalizes incoherent transformation within a neighborhood. It keeps points within a neighborhood relatively close after transformation and points far apart remain distant. Let $r_{ij}$ be the probability that two points $y_i$ and $y_j$ are neighbors before the transformation and $s_{ij}$ be the probability that they remain neighbors after the transformation. A constraint on the local structure is represented as the minimization of the cost function which is the sum of Kullback-Leibler (KL) divergences between $r_{ij}$ and $s_{ij}$ distributions over neighbors of each point:

$$\sum_i D_{KL}\left(\mathbf{R}_i \| \mathbf{S}_i\right) = \sum_{ij} r_{ij} \log \frac{r_{ij}}{s_{ij}}, \tag{13}$$

where

$$r_{ij} = \frac{\exp(-\beta_2 \left\| \mathbf{y}_i - \mathbf{y}_j \right\|_2^2)}{\sum_{k \neq i} \exp(-\beta_2 \left\| \mathbf{y}_i - \mathbf{y}_k \right\|_2^2)} \;,\; s_{ij} = \frac{\exp(-\left\| f(\mathbf{y}_i) - f(\mathbf{y}_j) \right\|_2^2)}{\sum_{k \neq i} \exp(-\left\| f(\mathbf{y}_i) - f(\mathbf{y}_k) \right\|_2^2)},$$

$\beta_2$ is precision parameter, and $\mathbf{R}_i = [r_{i1}, ..., r_{iM}]$ and $\mathbf{S}_i = [s_{i1}, ..., s_{iM}]$ are probability distributions.

The preservation of the global shape is achieved by enforcing landmark correspondences, which minimizes the total distance between paired key points:

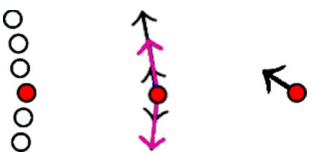

**Fig. 2** The local structure descriptor of a point (depicted as the solid red circle) is computed from its neighbors (depicted as circles). Vectors originating from the point to their neighbors (shown in the middle) are summed to get the local structure descriptor

$$\sum_{m,n}^{M,N} A_{m,n} \|\mathbf{x}_n - f(\mathbf{y}_m)\|_2^2, \tag{14}$$

where $\mathbf{A}_{M\times N}$ is landmark coefficient matrix, $A_{m,n} = 1$ if $(\mathbf{x}_n, \mathbf{y}_m) \in L$; otherwise 0, and $L$ is a set containing all pairs of landmark correspondences.

One way of regularizing the transformation function is minimizing the integral of the square of second derivatives of the spline function, i.e., penalizing large *curvature*. This smoothing constraint can be generalized to higher dimension (Hastie et al. 2009; Wahba 1990). For example, TPS-RPM (Chui and Rangarajan 2000b, 2003) employs the following regularization in the objective function:

$$\int\int \left[ \left( \frac{\partial^2 f}{\partial \mathbf{x}^2} \right)^2 + 2 \left( \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{y}} \right)^2 + \left( \frac{\partial^2 f}{\partial \mathbf{y}^2} \right)^2 \right] d\mathbf{x}\, d\mathbf{y}. \tag{15}$$

Table 2 lists the regularization terms used in point set registration methods.

The availability of large data sets (Bednarik et al. 2018; Bogo et al. 2017; Li et al. 2017) with ground truth and the development of deep networks enable the development of learning-based non-rigid point set registration methods (Li et al. 2019; Shimada et al. 2019; Wang and Fang 2019). Shimada et al. (2019) introduced the Displacements on Voxels Networks (DispVoxNets) that learns a deformation model and priors for different categories of point sets and regresses the displacement function on the sampled voxel grids. Wang and Fang (2019) presented an unsupervised coherent point drift network (CPD-Net) that learns the geometric transformation for an efficient non-rigid point set registration. Global feature descriptors are learned from the point sets and the drifts are predicted for each point. The network was trained in an unsupervised manner by minimizing the alignment loss function between the transformed source point set and the

**Table 2** Regularization terms used in registration methods

| Regularization | Method |
|---|---|
| $\int\int \left[ \left( \frac{\partial^2 f}{\partial \mathbf{x}^2} \right)^2 + 2 \left( \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{y}} \right)^2 + \left( \frac{\partial^2 f}{\partial \mathbf{y}^2} \right)^2 \right] d\mathbf{x}\, d\mathbf{y}$ | Chui and Rangarajan (TPS-RPM) (2000b, 2003), Chui and Rangarajan (MPM) (2000a), Feng and Feng (2020), Yang et al. (2018), Wang et al. (2008), Yang et al. (GLMD) (2015) |
| $\int \sum_{m=0}^{\infty} c_m (D^m f(x))\, dx$ | Myronenko and Song (CPD) (2010); Myronenko et al. (2007), Ge and Fan (GLTP) (2019); Ge et al. (2014), Ge and Fan (LSP) (2015), Calvo et al. (CCPD) (2018), Zho et al. (2019b), Zhang et al. (SCCD)(2017a), Song and Fan (2015), Zhou et al. (2014), Maharjan and Yuan (2020); Maharjan et al. (2021), Zhou et al. (DSMM) (2017) |
| $\sum_{j=1}^{M} \left\| f(\mathbf{y}_j, \theta) - \sum_{\mathbf{y}_k \in \tilde{\mathbf{Y}}_j} w_{jk} f(\mathbf{y}_k, \theta) \right\|_2^2$ | Ge and Fan (GLTP) (2019); Ge et al. (2014), Ge and Fan (LSP) (2015), Song and Fan (2015), Wang et al. (2018), Zho et al. (2019b) |
| $\sum_{m=1}^{M} \left\| \mathcal{C}(y_m) - \mathcal{C}(f(y_m, \theta)) \right\|_2^2$ | Ge and Fan (LSP) (2015), Zho et al. (2019b) |
| $\frac{1}{(u+l)^2} \sum_{i,j=1}^{l+u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$ | Ma et al. (MR-RPM) (2019a), Wang et al. (SCGF) (2017), Wang et al. (MRCVF) (2016a), Panaganti and Aravind (2015) |
| $\sum_{m,n}^{M,N} A_{m,n} \|\mathbf{x}_n - f(\mathbf{y}_m)\|_2^2$ | Maharjan and Yuan (2020); Maharjan et al. (2021) |
| $\sum_{ij} r_{ij} \log \frac{r_{ij}}{s_{ij}}$ | Maharjan and Yuan (2020); Maharjan et al. (2021) |

target point set. Li et al. (2019) developed an unsupervised point correspondence network (PC-NET). The method learns global shape descriptors with an encoder to capture the global and deformation-insensitive geometric properties. These descriptors are used to guide progressively deform the template shape toward the target shape in a motion-driven process. Loss is computed from the reconstructed and the original point sets.

### 2.1.3 Graph matching

Given the irregularly spaced points in a point set, the graph is a natural structure for point organization and processing. A number of methods used graph matching (GM) for establishing correspondences between point sets (Chang et al. 2020; Cour et al. 2007; Torresani et al. 2008, 2013). Given two graphs $G$ and $G'$, the point correspondence matrix $\mathbf{C} \in \{0,1\}^{M\times N}$ maximizes the following criterion:

$$\sum_{i,k} a_{ik}c_{ik} + \sum_{i,j,k,l} b_{ij,kl}\, c_{ik}c_{jl}, \tag{16}$$

such that

$$\mathbf{C}\mathbf{1}_N \le \mathbf{1}_M, \quad \mathbf{C}^\top \mathbf{1}_M \le \mathbf{1}_N \tag{17}$$

where $a_{ik}$ gives the similarity between *ith* vertex of $G$ and *kth* vertex of $G'$, $b_{ij,kl}$ gives the similarity between edge $(i, j)$ in $G$ and edge $(k, l)$ in $G'$, and $\mathbf{1}_M$ denotes a vector of $M$ ones. This objective function can be expressed in the form of Lawler's quadratic assignment problem (QAP) (Loiola et al. 2007):

$$E(\mathbf{C}) = \mathrm{vec}(\mathbf{C})^\top \mathbf{K}\mathrm{vec}(\mathbf{C}), \tag{18}$$

where $\mathrm{vec}(\mathbf{C}) \in \{0,1\}^{MN\times 1}$ denotes the vectorization of $\mathbf{C}$, and $\mathbf{K} \in \mathbb{R}^{MN\times MN}$ is an affinity matrix computed by:

$$\mathbf{K}_{(c_{ik},c_{jl})} = \begin{cases} a_{ij} & \text{if } i = j \text{ and } k = l, \\ b_{ij,kl} & \text{if } (i,j) \in E, (k,l) \in E', \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

Leordeanu and Hebert (2005) used GM to find point correspondences that employed a relaxed one-to-one mapping constraint [Eq. (17)] and integral constraint (i.e., element of the correspondence matrix can use real values in the range [0, 1] instead of binary values $\{0,1\}$, so $c_{ij} \in [0,1]$). The method builds a weighted adjacency matrix of a graph whose vertices are the potential correspondences and the weights between the vertices measure the pairwise agreement between the pair of correspondences. The key idea is that a strongly connected cluster is formed by the agreement links among the correct correspondences. By using the principal eigenvector of the adjacency matrix, the corrected correspondences are recovered, which are strongly associated with the main cluster.Torresani et al. (2008, 2013) extracted feature points from images and limited the matching to one correspondence at most per feature point. The objective function includes feature difference, unmatched features, geometric compatibility between correspondences, and spatial proximity of the matched features. GM was used to minimize the total similarity between vertices and edges.

Chang et al. (2020) grouped points into clusters and applied GM to get the correspondence between point groups, which achieved an initial coarse alignment for further refinement. Duchenne et al. (2009, 2011) applied higher-order GM techniques to find correspondences between feature sets, and the tensor power iteration technique was used to solve the higher-order matching problem. Zeng et al. (2016, 2010) applied the higher-order GM to get a sparse correspondence. Feature points were identified such as the local maxima of Gaussian curvature (Lipman and Funkhouser 2009) and the average geodesic distance function (Kim et al. 2011) on the input surfaces. These sparse but reliable feature points are matched via higher-order GM. In the refinement stage, the method employed a deformation model and used the sparse correspondence as a constraint to register the rest of the points.

The data structure, as well as the well-developed graph theory, make the GM-based method a proper choice for registering point sets. However, the computational complexity limits its usage to processing a relatively small set of points, e.g., feature points extracted from images or abstracted groups of points. Hence, it is often employed to generate a coarse matching for further refinement.

### 2.2 Deformation

Although the aforementioned non-rigid point set registration puts points into alignment, it is unquestionable that large deformation is still a great challenge that requires further investigation. In many real-world applications, we face the need to register point sets that deform greatly and unevenly (Ye et al. 2016). Figure 3 illustrates six point sets of human pose with large and uneven deformation. As shown in Fig. 3a–c, large deformation is in the upper body and the legs remain mostly unchanged. Deformation in such applications often varies substantially among parts of the body.

The study of dealing with large and uneven deformation in non-rigid point-set registration is very challenging due to several factors. For example, some regions of the point set might need large deformation while the other regions of the point set might need very little deformation. During the deformation, especially in a large case, the methods should maintain the local neighborhood geometry so that the structure of the local regions and the overall shape of the point set is maintained even in the case of large deformation. Also (and most importantly), methods need to avoid getting trapped into local minimum solutions while dealing with large and uneven deformation. This can usually happen when the

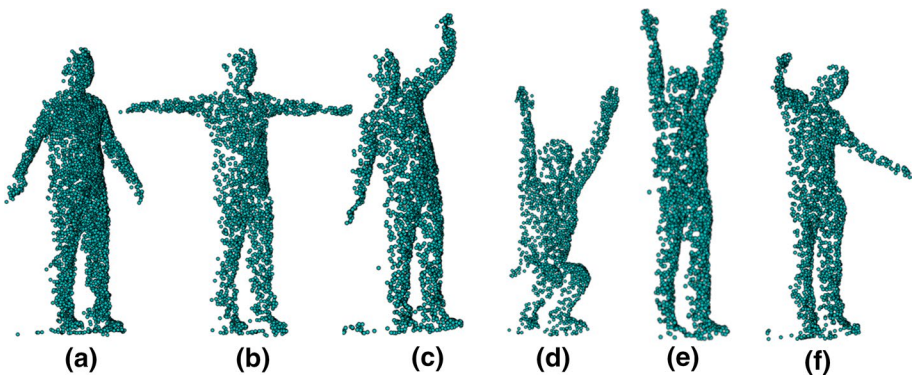


**Fig. 3** Point sets with large and uneven deformations (Maharjan et al. 2021)

method does not provide any prior clue or knowledge about the different corresponding regions and incorrect regions of the two point sets may be aligned. Figure 3 present one such case where large and uneven deformation is required to correctly register the point sets representing complex human poses (Maharjan et al. 2021). If a method needs to register, for example, both hands down pose (Fig. 3a) with both hands up pose (Fig. 3e), regions from the head, torso, and legs regions require relatively small deformation while hands regions need significant deformation to correctly register both point sets. In practice, large and uneven deformation during the registration is required in many real-world applications such as human pose estimation and tracking for physical training applications, gaming, etc. Thus far, existing registration methods have conducted studies on non-rigid point set registration to cope with large and uneven deformation. Here, we present different strategies in the existing methods to deal with it.

One strategy to deal with large deformation is to maintain local neighborhood structures during the transformation. For example, to handle highly articulated deformation such as 3D human data, (Ge and Fan 2019; Ge et al. 2014) extended the CPD by adding an LLE regularization term to maintain local neighborhood structure during transformation. The aim was to balance the CPD and LLE terms to handle complex deformations. The method needed careful selection of template point set (such as human T-pose, similar to Fig. 3b) to avoid local minima solutions. Later, the LSP method (Ge and Fan 2015) extended the GLTP and included an additional constraint to encode the local neighborhood scale based on Laplacian coordinate to deal with complex non-rigid and articulated deformations. However, the LSP, similar to the GLTP, is prone to get trapped into the local minimum solution in case of significant deformation, even though the local neighborhood structure is maintained (Maharjan et al. 2021). Tajdari et al. (2022) proposed semi-curvature as part of the cost function that also includes distance and stiffness for establishing the correspondences. The semi-curvature helps eliminate the conflations and represents the intrinsic properties across surfaces.

Another strategy typically includes two steps where highly reliable correspondences (usually a sparse set) are established first and then the final registration is completed with the help of the initial correspondences (Chang et al. 2020; Zeng et al. 2016). Zeng et al. (2016) used higher-order graph matching under the assumption of isometric deformations for sparse correspondences of the feature points of the source point set. These sparse correspondences were used to limit the search space of matching candidates for each point of the source point set to find accurate dense correspondences. Chang et al. (2020) relied on coarse-to-fine steps to deal with large deformation. The point clouds were divided into local groups and the correspondences between such groups were established by the graph matching technique. The initial alignment is followed by a fine non-rigid registration for accurate alignment.

Using prior knowledge of correspondences between salient regions has been used to deal with significantly large and uneven deformation in Maharjan and Yuan (2020); Maharjan et al. (2021). The salient regions typically represent unique entities such as the head, hands, and feet in a 3D human point set. Landmarks (or key points) which represent the salient regions are matched so that the point sets are aligned. Together with the landmark constraint and preserving local neighborhood structure preservation, the method handled the large and uneven deformation and avoided local minimum solution.

Recent learning-based non-rigid registration methods have made substantial progress to deal with the deformation issue. The learning-based methods usually consist of a feature extraction component followed by a correspondence search. Trappolini et al. (2021) adopt the transformer architecture as a geometrical translator between point clouds. Attention to

the underlying density of the geometry is used that takes a template point cloud and its properties (e.g., skinning weights or animation cues) as input and modifies its geometry to fit the target point cloud. Zeng et al. (2021) proposed an end-to-end deep learning method using deformation-like reconstruction. The network consists of a feature embedding module, a correspondence indicator, and a symmetric deformer, which learns the pointwise features to generate a learnable correspondence matrix to permute the input pair. Li and Harada (2022) integrate geometry features with the 3D positional information via position encoding. Non-rigid iterative closest point method is employed for the matching process. Feng et al. (2021) represent the non-rigid transformation with a combination of rigid transformations and perform registration iteratively with a recurrent network. The loss function measures the 3D shape similarity on the projected 2D views. Luo et al. (2022) integrate local geometry and global structure into a deep graph matching network. Graphs are constructed by taking high-dimensional features as nodes and point-wise Euclidean distance as a weight to edges. The topological structure is embedded into nodes using a graph embedding layer. To estimate the correspondence, an affinity layer and a sinkhorn layer are used.

Incorporating multiple information while computing the correspondences limits the search space of matching candidates for the points and hence helps to deal with large deformation in registration. For example, Feng and Feng (2020) fused both the spatial location of a point and the local neighborhood structure similarity around the point to compute reliable correspondences. Typically, feature descriptors such as shape context (Belongie et al. 2002) are employed for matching local neighborhood structures. However, it should be noted that the matching via feature descriptors are not always reliable, and is sensitive to noises and repeatable patterns/regions in the point sets.

In summary, identifying and incorporating landmark correspondence is an effective means to deal with the registration of point sets with large and uneven deformation. Given the extensive research on feature points, extraction, and methods for finding landmarks, efficiently establishing correct correspondences between landmark points is still a technical obstacle and requires further studies. The huge search space of correspondence and the inconsistent number of landmarks make it a non-trivial task. In addition, studies to explore computationally efficient and reliable landmark correspondences for various point sets are needed.

### 2.3 Data degradation

Data degradation such as noises, outliers, and occlusions is a major challenge in non-rigid point set registration. Many registration methods deal with the data degradation by preprocessing the data such as removing outliers. In reality, data degradation is unavoidable and it is imperative for registration methods to handle it carefully for accurate registration results. Figure 4 shows an example of data degradation in 3D face point sets (Myronenko and Song 2010). The template (source) point set (left) and target point set (right) both contain outliers but the target point set is noisy as well. Outliers (red dots in Fig. 4) are the extra points that do not represent any region or object and do not belong in the point set. Noisy points are the points that are distorted or perturbed. Due to occlusions or incomplete point sets, every point in a point set may not have its true correspondence in another point set (i.e., missing correspondence). In this case, points from a point set should not align with points in another point set. To minimize the effects of data degradation, existing methods use techniques such as regularizing the transformation functions, incorporating prior probabilities for points, and preserving local neighborhood geometry.

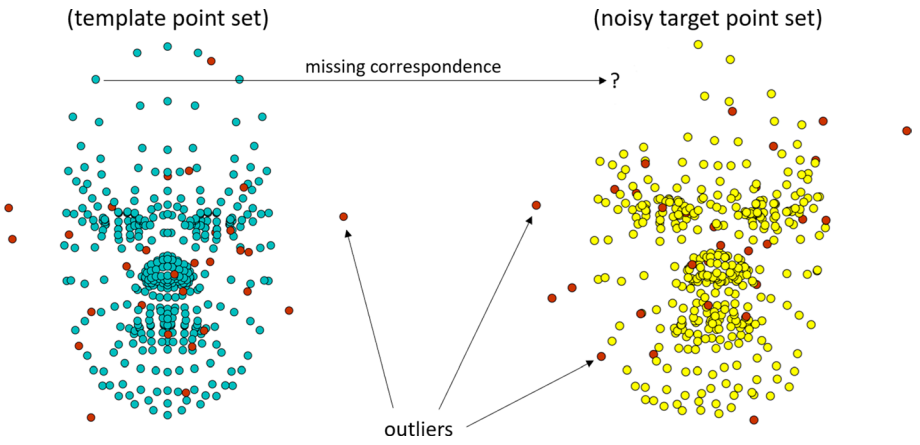


**Fig. 4** An example of data degradation in 3D face (Myronenko and Song 2010): template (source) point set (left) and noisy target point set (right). Outliers are shown in red circle dots in both point sets. A missing correspondence for a point in the template point set showing by an arrow at the top

One popular approach to deal with data degradation is to reject incorrect correspondences (Ma et al. 2015, 2019b, 2014; Wang et al. 2018, 2016a, 2017). In this approach, the putative correspondence set is computed and the true correspondences (inliers) are identified by rejecting the incorrect ones. The rejection criteria include the distance between the correspondence pair must be *less* than a threshold and the correspondence weight (probability of correspondence pair) is *greater* than a threshold. In general, putative correspondence set is computed using feature descriptors such as SIFT (Lowe 1999), Shape Context (SC) (Belongie et al. 2002), SURF (Bay et al. 2006), FPFH (Rusu et al. 2009), Spin Image (Johnson and Hebert 1999), IDSC (Ling and Jacobs 2007), 3DSC (Frome et al. 2004), etc. To compute this set, the feature descriptor for each point is computed and then the initial correspondence set is obtained using techniques such as computing the cost matrix using the chi-squared test and using the Hungarian method (Papadimitriou and Steiglitz 1982), sample consensus alignment (Rusu and Cousins 2011), and computing correlation coefficient for initial correspondence list.

Ma et al. (2015, 2013) used $L_2E$ estimator to improve robustness to noise and outliers. The initial putative correspondence set is computed using feature descriptors (SC (Belongie et al. 2002) in 2D and Spin Image (Johnson and Hebert 1999) in 3D). In their method, a correspondence pair $(\mathbf{x}_i, \mathbf{y}_i)$ is identified as true correspondence if the exponential distance is greater than a threshold. Wang et al. (2017) regularized the transformation function by employing smoothness and manifold regularization to reduce noises and outliers. The proposed spatially constrained Gaussian fields (SCGF) method extracts the first feature descriptor using shape context information (IDSC (Ling and Jacobs 2007) for 2D and 3DSC (Frome et al. 2004) for 3D) and the chi-squared test is used to compute the cost matrix, which is used in the Hungarian method to obtain one-to-one correspondence. The true correspondence set includes only the pairs with a large exponential distance. Ma et al. (2014) relied on local feature descriptors (such as shape context (Belongie et al. 2002) for 2D and MeshHOG (Zaharescu et al. 2009) for 3D) to create putative correspondences and derive correspondence by interpolating a smooth vector field between the point sets. The inlier set is obtained by collecting pairs that have a large corresponding probability. Later, the VFC method is extended

with manifold regularization (Belkin et al. 2006) in the transformation function to deal with incorrect correspondences in Ma et al. (2019a); Wang et al. (2016a). Ma et al. (2017, 2019b) assumed that local neighborhood structures are maintained between true correspondences. The initial correspondence list was computed between sets using feature descriptors (e.g., SIFT Lowe 1999). Correct correspondences are identified as those with small local neighborhood structure differences. Gaussian Field Consensus (GFC) was proposed to reject outliers given a putative correspondence set in Wang et al. (2018), in which the topological structure of the feature points is maintained and Gaussian distances between the points are used. The correspondence is identified as those with large correspondence weight. Wang et al. (2022) formulated the registration problem as a partial distribution matching process, which avoids exact point-wise matching and achieves robustness against outliers. The method leverages a partial Wasserstein adversarial network to approximate the discrepancy using a neural network and incorporates a coherence regularizer for smooth transformation.

In GMM models, prior probabilities are the same for all centroids of the mixture model (Ge and Fan 2015; Myronenko and Song 2010), whereas the posterior probabilities of the input points to the centroids are different. Instead of using equal prior probabilities, recent methods assigned or dynamically computed prior probabilities to deal with data degradations (Ma et al. 2016; Panaganti and Aravind 2015; Tao and Sun 2014; Zhang et al. 2017a). The idea is to compute the similarities between the points and the centroids. However, the prior probabilities computed with the help of feature descriptors are not reliable and the methods also need to rely on regularization terms such as motion coherence (Yuille and Grzywacz 1988) and manifold regularization (Belkin et al. 2006) to identify correct correspondences. Wang and Chen (2021) employ locally linear embedding to represent the topological structure of location neighborhoods and apply context-aware Gaussian fields for spatial coherent point matching. The results on point sets with affine transformation demonstrate improved robustness to outliers. Chen et al. (2022) used Hausdorff distance to measure the similarity of the local structure of the point set. The transformation between the two-point sets is determined by the reproducing kernel Hilbert space based on motion coherence. Cao et al. (2022) proposed a probabilistic model that includes an outlier component based on the asymmetric generalized Gaussian mixture model that aggregates unevenly distributed outliers for elimination.

The registration methods based on density estimation using mixture model add a uniform distribution, $p_u = \frac{1}{N}$, to account for outliers and noises (Chui and Rangarajan 2000a; Ge et al. 2014; Maharjan et al. 2021; Myronenko and Song 2010). The weight of this uniform distribution, namely the outlier ratio ($\gamma$), is manually specified

$$p(\mathbf{x}_i) = (1-\gamma)\sum_{j=1}^{M} p(\mathbf{x}_i|\mathbf{y}_j)p(\mathbf{y}_j) + \gamma p_u. \tag{20}$$

Zhang et al. (2017a) dynamically computed the outlier ratio based on rotation invariant shape context (RISC). The RISC feature descriptor is based on shape context (Belongie et al. 2002) but the main reference x-axis is computed using the principal component analysis (PCA) (i.e., the eigenvector corresponding to the largest eigenvalue of the covariance matrix of the point set), which makes it rotation invariant. To compute the outlier ratio, RISC is constructed for every point. The point from $\mathbf{x}_i \in \mathbf{X}$ is derived from the centroid $\mathbf{y}_j \in \mathbf{Y}$ if the RISC matching has the highest value between them. The ratio between the

number of points not derived from GMM and the total number of points in $\mathbf{Y}$ is used as the outlier ratio.

Using only one additional uniform distribution represented by a single Gaussian component (see Eq. (20)) in the density estimation models likely degrades the registration results in the case of heteroscedastic distributed outliers. Qu et al. (2017) applied two Gaussian mixture models to estimate heteroscedastic noise and outliers during transformation and correspondence estimations.

### 2.4 Data integration

A widely used technique in the existing methods is to integrate prior correspondence knowledge or information between point sets (Tao and Sun 2014; Wang and Chen 2017; Yang et al. 2018; Zhang et al. 2017a; Zhou et al. 2018). The prior knowledge is encoded as a correspondence weight or probability and is integrated as either soft-assignment or binary assignment by taking the advantages of feature descriptors (Belongie et al. 2002; Ling and Jacobs 2007; Lowe 1999; Rusu et al. 2009). The feature descriptors encode the neighborhood topological structure and assume a similar neighborhood structure of two corresponding points. In reality, these prior correspondences are not always accurate, which motivates the development of correspondence rejection techniques (Ma et al. 2014; Wang et al. 2016a, 2018). Nonetheless, the prior correspondences help the registration methods to deal with outliers, and noises and preserve the local neighborhood structure. To compute the prior probability, methods typically relied on histogram-based feature descriptors such as shape context (Belongie et al. 2002) in 2D and Fast Point Feature Histogram (Rusu et al. 2009) in 3D. The feature descriptors are obtained for each point from both point sets and the chi-squared test is used to compute the cost matrix

$$\pi_{ij} = \sum_{h=1}^{H} \frac{(F_i(h) - F_j(h))^2}{F_i(h) - F_j(h)}, \tag{21}$$

where $F_k$ is a histogram-based feature descriptor and $h$ is a bin number. The Hungarian algorithm is used for binary assignment (Ma et al. 2016; Zho et al. 2019b) or the following exponential function is used to compute soft-assignment for prior probability (Panaganti and Aravind 2015; Wang et al. 2016a)

$$w_{ij} = \frac{\exp(-\alpha\pi_{ij})}{\sum_{m=1}^{M} \exp(-\alpha\pi_{im})}. \tag{22}$$

Golyanik et al. (2016) incorporated a correspondence prior based on CPD. The initial correspondences are obtained by comparing feature descriptors such as Persistent Feature Histograms (PFH) (Rusu et al. 2008) at Intrinsic Shape Signature descriptor (ISS) (Zhong 2009) 3D keypoints.Zhou et al. (2018) used Dirichlet distribution to model the prior probability between a point and a Student's T mixture model component to overcome underfitting and improve robustness to data degradation. Typically, prior correspondence is computed in every iteration during the optimization process, so these methods could be computationally expensive if the size of the point sets is large.

A new deformable point set registration framework is proposed to incorporate constraints in the deformation model in Kolesov et al. (2013a, 2013b, 2016). In the framework, the deformation model is defined by an additive combination of a rigid transformation and

a set of radial basis functions. The transformation function contains linear ($f_{lin}$) and non-linear ($f_{nl}$) terms as follows:

$$f(\mathbf{y}_m, \theta) = f_{lin}(\mathbf{y}_m, \theta) + f_{nl}(\mathbf{y}_m, \theta), \tag{23}$$

where

$$f_{lin}(\mathbf{y}_m, \theta) = \mathbf{R}\mathbf{y}_m + \mathbf{t} \ , \ f_{nl}(\mathbf{y}_m, \theta) = \sum_{i=1}^{G} w_i \phi_g(\mathbf{y}_m, \theta),$$

$\mathbf{R}$ is the linear component, $\mathbf{t}$ is translation vector, $\phi_g(,)$ is Gaussian radial basis function. To ensure orientation preservation and injectivity constraints, the following constraint is employed:

$$\det(J(f(\mathbf{x}, \theta))) > 0, \forall \mathbf{x} \in S, \tag{24}$$

where $J(.)$ returns a Jacobian matrix, det(.) computes the determinant, $S \subset \mathbb{R}^d$ is an open subset containing the region of interest. Hirose (2017) applied a PCA-based statistical shape model to encode prior shapes for registering the mean shape and point set. In this method, the statistical shape model is combined with the similarity transformation and motion coherence for the transformation model. Given the mean shape $\mathbf{u} = (u_1^T, ..., u_M^T)^T \in \mathbb{R}^{MD}$, we have $K$ shape variations $\mathbf{H} = (\mathbf{h}_1, ..., \mathbf{h}_k) \in \mathbb{R}^{MD \times K}$ and $H_m \in \mathbb{R}^{D \times K}$ is a submatrix of $\mathbf{H}$ that corresponds to the $m^{th}$ landmark $u_m \in \mathbb{R}^D$ in the mean shape $\mathbf{u}$. The transformation model is defined as,

$$f(u_m, \theta) = sR(u_m + H_m z) + t, \tag{25}$$

where $s \in \mathbb{R}$ is a scale factor, $R \in \mathbb{R}^{D \times D}$ is a rotation matrix, $t \in \mathbb{R}^D$ is a translation vector, and $z = (z_1, ..., z_K)^T \in \mathbb{R}^K$ is a weight vector. This technique deals with different categories of point sets but requires training data to compute $K$ shape variations matrix $\mathbf{H}$ for each type of point set. Another approach is to include information such as color or point normals in the registration methods (Saval-Calvo et al. 2018) or human motion tracking (Horaud et al. 2009). Color Coherent Point Drift (CCPD) extended CPD and incorporated both color and spatial location of points for establishing correspondences between the point sets (Saval-Calvo et al. 2018).

As many low-cost imaging instruments including autonomous vehicles and mobile devices capture depth images, the variety of the modality and data density make registration much more difficult when data captured with different devices need to be aligned. Huang et al. (2017) explored cross-source point cloud registration in a rigid case. While others studied group-wise registration of the point sets, which are obtained from similar devices (Wang et al. 2008). Investigations on registering multiple point sets acquired with different devices are highly desired.

### 2.5 Computational efficiency

Computational efficiency is one of the main challenges in many existing non-rigid point set registration methods. Several important factors that make the registration methods inefficient include computation of affinity matrices which involve all pairs of points between the point sets, computational costly matrix multiplications in the point sets containing a

large number of points, and iterative process to find the optimal set of the parameter during optimization which is necessary and employed in most of the existing methods. Many studies (Combes and Prima 2020; Hirose 2021; Lian et al. 2017) have been conducted to tackle this computational efficiency issue using techniques such as Fast Gaussian Transform (FGT) (Greengard and Strain 1991), downsampling, and interpolation (Hirose 2020), and learning-based methods.

Myronenko and Song (2010) applied fast Gauss transform (FGT) (Greengard and Strain 1991) and Low-Rank matrix approximation techniques to handle point sets with a large number of points. In correspondence models based on soft-assigned, especially in density estimation methods (Maharjan et al. 2021; Myronenko and Song 2010), a computationally expensive sum of exponentials is required. FGT computes the sum of exponentials as follows

$$F_G(\mathbf{y}_m) = \sum_{n=1}^{N} v_n \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_n - \mathbf{y}_m\|^2\right), \quad \forall \mathbf{y}_m, m = 1, \ldots, M, \tag{26}$$

where $\mathbf{v} = [v_1, \ldots, v_n]^T$ is a vector. It takes $O(MN)$ for direct evaluation of Eq. (26) but takes $O(M + N)$ using the FGT. Another bottleneck in many registration methods is the computation of Gaussian affinity matrix $\mathbf{G}$. To reduce the computational cost, this matrix $\mathbf{G}$ is approximated using a low-rank approximation, $\tilde{\mathbf{G}} = \mathbf{V}\mathbf{D}\mathbf{V}^T$, where $\mathbf{D}_{K\times K}$ denotes a diagonal matrix with $K$ largest eigenvalues and $\mathbf{V}_{M\times K}$ is formed from the corresponding eigenvectors. Using matrices $\mathbf{V}$ and $\mathbf{D}$ in the EM iterations, instead of $\mathbf{G}$, greatly reduced the computational costs (Myronenko and Song 2010). Golyanik et al. (2016) also used the FGT technique to reduce computational costs but switched to the truncated Gaussian approximation (set zeros outside a predefined box) when the covariance of the Gaussians becomes small in the final iterations. Dupej et al. (2015) pointed out that the low-rank approximation of large Gaussian affinity matrix $\mathbf{G}$ before the EM iterations are still computationally expensive and they improved the low-rank approximation of $\mathbf{G}$ based on Nyström method (Williams and Seeger 2001) and improved fast Gauss transform (IFGT) (Yang et al. 2003). The downside of using these techniques is that the registration accuracy may degrade.

Recently, Hirose (2021) presented Bayesian CPD (BCPD), which formulates CPD in the Bayesian framework. Nyström method (Williams and Seeger 2001) is used to accelerate the computation of the Gaussian affinity matrix. An accelerated method BCPD++ down-samples the point sets and uses the interpolation for an efficient registration (Hirose 2020). The downsampled point sets are registered using BCPD and the displacement vectors for the removed points during the downsampling process are interpolated. This allows BCPD++ to handle a large number of points.

One of the main benefits of using learning-based methods, especially DNN-based methods, is to avoid iterative processes for non-rigid point set registration. The end-to-end learning-based methods formulate image (or point set) registration as a regression or a classification problem (Pais et al. 2020). The model derives a transformation matrix that aligns two inputs, and these methods mostly deal with rigid registration problems (Huang et al. 2020; Lu et al. 2019c). The existing DNN-based methods, whether they are supervised (Shimada et al. 2019) or unsupervised (Li et al. 2019; Wang and Fang 2019), however, need a very large amount of training data and take high computational power to generate a model. Once the parameters are learned from training the networks, which is usually a computationally expensive process, the DNN-based methods are typically many magnitudes faster than the existing iterative methods. Yet, a side effect is poor generalization performance to different

**Table 3** Registration time in seconds

| Method | Data set | # of points | Time |
|---|---|---|---|
| Feng and Feng (2020) | Fish (Chui and Rangarajan 2003) | 98 | 1.7 |
| | IMM Face (Nordstrøm et al. 2004) | 58 | 0.7 |
| | IMM Hand (Stegmann and Gomez 2002) | 56 | 0.7 |
| | Complicated Chinese Char (CCC: complicated chinese characters, https://github.com/xdregis/complicated_chinese_characters) | 161 | 4.4* |
| Zho et al. (2019b) | Fish (Chui and Rangarajan 2003) | 98 | 2.5 |
| | Chinese character (Chui and Rangarajan 2003) | 100 | 1.8 |
| | IMM Hand (Stegmann and Gomez 2002) | 56 | 0.5 |
| Qu et al. (2017) | Chinese character (Chui and Rangarajan 2003) | 105 | 4.2 |
| | Road | 277 | 6.2 |
| Wang et al. (2018) | Oxford Buildings (Philbin et al. 2007) | – | 0.2* |
| Myronenko and Song (2010) | 3D Bunny (3DScan: 3D Scanning Library, http://graphics.stanford.edu/data/3Dscanrep) | 3504 | 3628.0* |
| (CPD, FGT+Low-Rank) | 3D Bunny (3DScan: 3D Scanning Library, http://graphics.stanford.edu/data/3Dscanrep) | 11,615 | 177.5* |
| Hirose (2020) (BCPD++) | Stanford 3D Scanning Repository (3DScan: 3D Scanning Library, http://graphics.stanford.edu/data/3Dscanrep) | 3.9M | 55.5* |
| Hirose (2021) (BCPD) | Stanford 3D Scanning Repository (3DScan: 3D Scanning Library, http://graphics.stanford.edu/data/3Dscanrep) | 514K | 1425.0* |

The ones with an asterisk * are the average time of multiple cases reported in the literature. The numbers are rounded to the tenth place

objects (or sometimes referred to as different categories). For example, a DNN-based non-rigid registration method trained with point sets of chairs performs poorly when tested with point sets of airplanes.

Table 3 presents the average time in seconds for completing registration using various data sets and the number of points used in the experiments. For point sets of small size (less than 100 points), the method by Feng and Feng (2020) took less than two seconds. It took about 3 to 5 s for slightly bigger point sets (130–190 points) from the Complicated Chinese Character data set (CCC: complicated chinese characters, https://github.com/xdregis/complicated_chinese_characters). Similarly, the method by Zho et al. (2019b) took under three seconds to complete registrations for smaller point sets such as Fish and Chinese Character (Chui and Rangarajan 2003), IMM Hand (Stegmann and Gomez 2002). The method by Qu et al. (2017) took four to six seconds to register points from 100-300 points. It is difficult and unfair to directly compare the results of these methods but all these methods did compare with the CPD in their experiments, which we use as the baseline for our discussion here. Typically, CPD takes less than one second to register smaller point sets mentioned above (Qu et al. 2017; Zho et al. 2019b) and the additional time these methods took to register is due to computational complexity introduced by incorporating additional constraints such as preserving local geometry, computing feature descriptors, etc.

The last two rows of Table 3 show the results of the methods handling much larger point sets. CPD (Myronenko and Song 2010) performed significantly faster when FGT and Low-Rank approximations were used. Without leveraging the approximation technique,

CPD was unable to register the Bunny point set with more than 35K points. Hence, the average time of the successful experiments is reported. The results suggest the benefits of using FGT and Low-Rank approximation techniques to deal with large point sets. Methods such as BCDP (Hirose 2021) and BCPD++ (Hirose 2020) handle a very large number of points, i.e., more than 100,000 points. As shown in this table, BCPD++ is much faster than BCPD. The registration of BCPD for the Lucy data set did not converge within 24 h (Hirose 2020), which was not reported and hence excluded from the average time. The results show a clear advantage of using down-sampling and interpolation techniques for dealing with very large point sets.

Registration efficiency is a bottleneck for applying the existing methods to real-world applications, especially when the size of the point sets is large. Downsampling is often used for improved efficiency, which, however, is far from effective to achieve real-time performance. BCPD++ (Hirose 2020) shows promising results based on downsampling and interpolation techniques in very large point sets but needs further explorations on significantly large and uneven deformation. Deep Neural Networks (DNNs) based methods also show encouraging results on the registration speed. Yet, difficulties faced by such methods include poor generalization to different categories, high training costs, and handling large point sets.

## 3 Data sets and evaluation

In this section, we present data sets used in the experiments and discuss commonly used evaluation techniques used in the existing methods. Preparation of data sets and evaluation is a difficult task, especially when the ground truth correspondences between the point sets are difficult to obtain. Therefore, many existing methods come up with a creative way of evaluating the registration methods.

### 3.1 Data sets

Table 4 presents publicly available data sets that have been used for the evaluations of point set registration methods. The data sets are categorized into two groups: with and without ground truth (i.e., correspondence among points). The data sets with ground truths are those that have either one-to-one correspondences between all the points or the landmarks of the point sets.

Typically, synthetic data sets such as Fish and Chinese character (Chui and Rangarajan 2000b, 2003) contain the ground truths. In some image data sets where one-to-one correspondences are impossible or expensive to prepare, a number of landmarks are identified and correspondences are established manually between the images such as in IMM Face (Nordstrøm et al. 2004), IMM Hand (Stegmann and Gomez 2002), WILLOW-ObjectClass (Cho et al. 2013). There are other cases where the one-to-one dense correspondences are established by other registration methods. For example, the SCAPE human data set (Anguelov et al. 2005) contains articulated human poses and the correspondences between the points of the human poses are established by correlation correspondence (Anguelov et al. 2005). The data sets are designed to evaluate specific aspects.

The last column of Table 4 presents the typical cases (but not limited to only those cases) where the data sets are used for. In contour registration, the point sets contain the contour (shape) of the objects (e.g., Fish or Chinese characters) and usually have fewer points. This

**Table 4** Publicly available data sets

| | Data set | | Dim. | # of cases | Avg. # points | Usage |
|---|---|---|---|---|---|---|
| With | G1 | Fish & Chinese character (Chui and Rangarajan 2003) | 2 | 4 | 100 | DD, CR |
| Ground | G2 | CMU sequence data (Datasets C Cmu house and hotel sequence images, http://vasc.ri.cmu.edu/idb/html/motion/house/index.html) | 2 | 212 | 30 (L) * | LR, TC |
| Truth | G3 | 2D Fish & 3D Face (Myronenko and Song 2010) | 2,3 | 52 | 392 | LR |
| | G4 | IMM Face (Nordstrøm et al. 2004) | 2 | 240 | 58 (L) * | LR |
| | G5 | DIR-LAB (4DCT and COPD) (Castillo, Dir-lab, https://www.dir-lab.com/index.html) | 3,4 | – | 300 | LR |
| | G6 | Open curve & closed contours (Chui and Rangarajan 2003) | 2 | 6 | 55 | DD, LR |
| | G7 | IMM Hand (Stegmann and Gomez 2002) | 2 | 40 | 56 (L) * | LR |
| | G8 | WILLOW-ObjectClass (Cho et al. 2013) | 2 | 305 | 10 (L) * | LR |
| | G9 | SHREC'19 (Marin et al. 2020) | 3 | 44 | 38,082 | D, SP |
| | G10 | The Oxford Buildings Dataset (Philbin et al. 2007) | 2 | 5062 | 11 * | TC |
| | G11 | The Paris Dataset (Philbin et al. 2008) | 2 | 6412 | – * | TC |
| | G12 | Complicated Chinese Character (CCC: complicated chinese characters, https://github.com/xdregis/complicated_chinese_characters) | 2 | 7500 | 161 | DD |
| | G13 | Space-time faces (Zeng et al. 2004) | 3 | 384 | 23,728 | D |
| | G14 | TOSCA (Bronstein et al. 2008a) | 3 | 80 | 50,000 | D |
| | G15 | SCAPE human (Anguelov et al. 2005) | 3 | 71 | 12,500 | D, SP |
| | G16 | FLAME (Li et al. 2017) | 3 | 46,905 | 5,023 | LR |
| | G17 | Texture-less Deformable Surfaces (Bednarik et al. 2018) | 3 | 26,445 | – | D, TC |
| | G18 | Horse, Camel and Elephant Sumner and Popović (2004) | 3 | 337 | 20,000 | D, TC |
| | G19 | ACCAD Mocap Dataset (Swagger) (OSU A Accad motion capture data, https://accad.osu.edu/research/motion-lab/mocap-system-and-data) | 3 | 581 | 126 (L) | D, LR |
| | G20 | Thin plate (Golyanik et al. 2018) | 3 | 4648 | – | D |
| | G21 | Dynamic FAUST (Bogo et al. 2017) | 4 | 7948 | 6,890 | D |
| Without | W1 | Stanford 3D Scanning Repository (3DScan: 3D Scanning Library, http://graphics.stanford.edu/data/3Dscanrep) | 3 | 379 | 77,734 | DD |
| Ground | W2 | Tools Dataset (Bronstein et al. 2008b) | 2 | 35 | 16,622 | CR, DD |
| Truth | W3 | Free-Viewpoint Video (FVV) (Lincoln) (Collet et al. 2015) | 3 | 508(F) | 1.60M | D |
| | W4 | Archive3D (Archive3D: Archive3d, http://archive3d.net/) | 3 | 55,632 | – | D |

The ones with an asterisk * are the data sets with landmarks. The symbols used in this table include: L (number of landmarks), F (number of frames), CR (Contour Registration), DD (Data Degradation), LR (Landmark Registration), TC (True Correspondence in Outlier Rejection), and D (Deformation)

case is very common and popular among many methods to evaluate the shape or outline of the point sets. Robustness to data degradation such as deformation, noise, outliers, rotation, and occlusion is one of the major challenges of the existing methods. Data sets such as Fish and Chinese character (Chui and Rangarajan 2000, 2003b) and Bunny from the Stanford 3D Scanning Repository (3DScan: 3D Scanning Library, http://graphics.stanford.edu/data/3Dscanrep) are frequently used data sets to evaluate the robustness with respect to data degradation. Landmark registration involves point sets that represent landmark points of the object and is used to evaluate the performance of methods for landmark-to-landmark correspondences. The landmarks are typically points such as corners of a house in Willow-ObjectClass (Cho et al. 2013) or points on a 3D face in IMM Face (Nordstrøm et al. 2004). Many existing methods identify the true correspondences from initial putative correspondences between the point sets. To evaluate the performance, existing methods use data sets with or without ground-truth correspondences. In the absence of ground truth correspondences, true correspondences are typically identified as those where the exponential distance (or correspondence weight) between the corresponding points is greater than a threshold. To evaluate large and uneven deformation, data sets including articulated human poses and shapes such as SHREC'19 (Marin et al. 2020), TOSCA (Bronstein et al. 2008a), SCAPE human (Anguelov et al. 2005) are used.

The data sets presented in Table 4 support the evaluation of new methods for research challenges discussed in Sect. 2. For evaluating the deformation, eleven data sets are suitable, including SHREC'19 (Marin et al. 2020), Space-time faces (Zeng et al. 2004), TOSCA (Bronstein et al. 2008a), SCAPE human (Anguelov et al. 2005), Texture-less Deformable Surfaces (Bednarik et al. 2018), Horse, Camel and Elephant (Sumner and Popović 2004), ACCAD Mocap Dataset (Swagger) (OSU A Accad motion capture data, https://accad.osu.edu/research/motion-lab/mocap-system-and-data), Thin plate (Golyanik et al. 2018), Dynamic FAUST (Bogo et al. 2017), Free-Viewpoint Video (FVV) (Lincoln) (Collet et al. 2015), and Archive3D (Archive3D: Archive3d, http://archive3d.net/). Most of these data sets are large and consist of ground truth correspondence, which enables comprehensive quantitative evaluation. For evaluating data degradation, Fish and Chinese character (Chui and Rangarajan 2003) are widely used. In addition, Open curve and closed contours (Chui and Rangarajan 2003), Complicated Chinese Character (CCC: complicated chinese characters, https://github.com/xdregis/complicated_chinese_characters), Stanford 3D Scanning Repository (3DScan: 3D Scanning Library, http://graphics.stanford.edu/data/3Dscanrep), and Tools Dataset (Bronstein et al. 2008b) provide more options for the evaluation. Besides these available data sets, techniques such as adding additive noise and outliers are often used in the evaluation. To understand the impact of outliers on the registration process, several data sets provide correspondence in outlier rejection such as CMU sequence data (House and Hotel) (Sumner and Popović 2004), The Oxford Buildings Dataset (Philbin et al. 2007), The Paris Dataset (Philbin et al. 2008), Horse, Camel and Elephant (Sumner and Popović 2004), and Texture-less Deformable Surfaces (Bednarik et al. 2018).

Point set registration is a key step in applications such as motion tracking, pose estimation, object classification/detection, and deformation analysis. The data sets included in Table 4 enable quantitative evaluation of general-purpose registration methods as well as methods for specific applications. TOSCA (Bronstein et al. 2008a), SCAPE human (Anguelov et al. 2005), ACCAD Mocap Dataset (Swagger) (OSU A Accad motion capture data, https://accad.osu.edu/research/motion-lab/mocap-system-and-data) contain point sets of complex human body poses, which support the evaluation of pose estimation. In addition, data sets with 4D data such as Dynamic FAUST (Bogo et al. 2017) that

contains a 3D body in motion with ground-truth correspondences are suitable for human motion analysis and tracking applications. Tracking applications can take advantage of data sets containing sequential data such as CMU sequence data (Datasets C Cmu house and hotel sequence images, http://vasc.ri.cmu.edu/idb/html/motion/house/index.html). There are data sets related to 2D or 3D facial expressions, including Space-time faces (Zeng et al. 2004), FLAME (Li et al. 2017), and IMM Face (Nordstrøm et al. 2004), which are useful for facial expression recognition and animation. Archive3D (Archive3D: Archive3d, http://archive3d.net/) contains a large number of 3D models of different objects such as sofas, doors, armchairs, tables, etc., which can be used for object classification and shape analysis. Another important collection of data set is related to medical applications, including DIR-LAB (4DCT and COPD) (Castillo, Dir-lab, https://www.dir-lab.com/index.html). The data set contains Thoracic 4DCT images and CT image pairs of inspiratory and expiratory breath-hold of lungs which are useful for deformation and functional analysis. The Oxford Buildings Dataset (Philbin et al. 2007) and the Paris Dataset (Philbin et al. 2008) have been used for evaluating the performance of object and landmark detections.

### 3.2 Evaluation techniques

Non-rigid registration methods are evaluated in different aspects. The most common evaluation aspects include robustness to data degradation, speed of the methods, and handling of large deformations. Existing methods use both qualitative and quantitative techniques to evaluate the performances in these aspects. Figure 5 summarizes the data sets and evaluation metrics used in the existing methods. The notation of data sets follows the indexes used in Table 4 that are grouped into with (G1–G21) and without ground truths (W1–W4). In the ones with ground truths, data set G1 [Fish and Chinese Character (Chui and Rangarajan 2003)] is used most often. The majority of the methods use more than one data set in the evaluation. Each evaluation metric is represented with a unique symbol as shown in the legend. Besides quantitative metrics, qualitative evaluation is also included, which is reported in all studies and hence is not marked throughout the table. However, for the data sets without ground truth, qualitative evaluation was only performed for some, e.g., data set W4. Both the root mean squared error (RMSE) and Euclidean distance have been widely employed in many studies. Among the 40 studies in this table, 18 studies employed Euclidean distance, and 17 used RMSE. RMSE and Euclidean distance are highly similar because most RMSE is computed based on the Euclidean distance. However, Euclidean distance is more plausible when ground truth is absent. Another popular metric is the matching rate, which is used to assess the correct identification of the corresponding points. Similar to accuracy, precision, and recall, the matching rate is an indirect evaluation measure for registration performance. The most often used metrics are highlighted with colored symbols in this figure.

Robustness to data degradation is an important aspect of a registration method and many methods include an evaluation of one or many scenarios of data degradation (Dan et al. 2018; Lian et al. 2017; Ma et al. 2016; Wang and Chen 2017; Wang et al. 2017). Data degradation typically includes deformations, noises, outliers, rotation, and occlusions. For each degradation type, experiments are conducted on multiple levels of degradation. For example, deformation levels from 1-5 are typically used and both qualitative and quantitative measures are presented. Here, level 1 means the minimum deformation and level 5 means the greatest deformation. In general, if different levels of deformed point sets are not present in the data sets, the points are perturbed to conduct the deformation experiments.

In the evaluation of the impact of noise, zero-mean white noise is often used with various standard deviations to the point coordinates. To mimic outliers, random points are added in proportion to the levels of distortion. Points from a certain region of the point set are removed explicitly for occlusion experiments. However, in data sets such as human pose point sets, self-occlusions are usually evaluated. Point sets are also often rotated with different degrees to evaluate the robustness of rotation.

Table 5 presents the quantitative registration performance of five methods with respect to data used in the evaluation. Many methods reported the experimental results using plots and visual illustrations and are excluded from this table. The first methods include the average RMSE error at different levels of deformation, outliers, and occlusions, which are presented in a triple. PR-GLS (Ma et al. 2016) reported lower RMSE errors for deformation, outliers, and occlusions in the 3D Wolf shape point set, and the method show robustness to data degradation. MR-RPM (Ma et al. 2019a) presented results using Fish and Chinese Character (Chui and Rangarajan 2003) point sets and achieved low errors for all the occlusion cases, that is, the method is less sensitive to occlusions. Note that it is difficult to compare the quantitative results because the data degradation levels reported in the literature are inconsistent even though the same point sets were used. In addition, it is an open problem where there is no standard or consistent way to obtain different levels of data degradation. Many existing methods either used the data sets that already contain levels of degradation or fabricated degradation by introducing noise or distortion to point coordinates.

The last three methods in Table 5 provide no differentiation among degradation and an average error is reported for each data set. The results of GL-CATE (Zeng et al. 2017b) suggest incorporating both global and local features achieves improved performances. The last two methods used Euclidean distance as the evaluation metric. Yet, the error of DSMM (Zhou et al. 2018) is the average physical Euclidean distance in millimeters (mm); whereas Yang et al. (2018) reported the average Euclidean distance between the corresponding points. Because different data sets and metrics were used in these methods, it is difficult to draw a conclusion about the performance. Even when the same metric was used, e.g., RMSE, because of the point spacing of the data set, the results could be drastically different. A normalized RMSE error using the average point spacing is helpful in future studies to provide a more plausible comparison.

Table 6 lists the performances in terms of the precision and recall of outlier rejection of methods using different data sets. The evaluation is formulated as a decision of correct correspondence. Wang et al. (2016a) used both synthetic data sets and data sets of images to evaluate MRCVF. The average precision and recall of MRCVF across all data sets are 98.03% and 98.37%, respectively. Using different data sets, Ma et al. (2014) reported an average precision and recall of 96.11% and 97.75%, respectively. Ma et al. (2017) achieved an improved precision of 99.16% and recall of 99.36% using the Affine Covariant Regions data set (Mikolajczyk et al. 2005). Leveraging the correct correspondence between point sets, the average Euclidean distance of VFC (Ma et al. 2014) and LPM (Ma et al. 2017) is in the range of [0, 0.04] with an improved performance by LPM when the occlusion and outlier are in a greater magnitude.

Robustness to data degradation such as noises, outliers, rotations, deformations, and occlusions is key to any successful registration method. Among the aforementioned data sets, few, if any, provide quantitative levels of degradation for each point set. Many methods attempted to provide an in-depth evaluation of the levels of data degradation to registration performance. However, the experiments present inconsistency among different studies. It is, hence, necessary to have the means to quantitatively evaluate the data degradation and include that into data sets for future studies and evaluation.

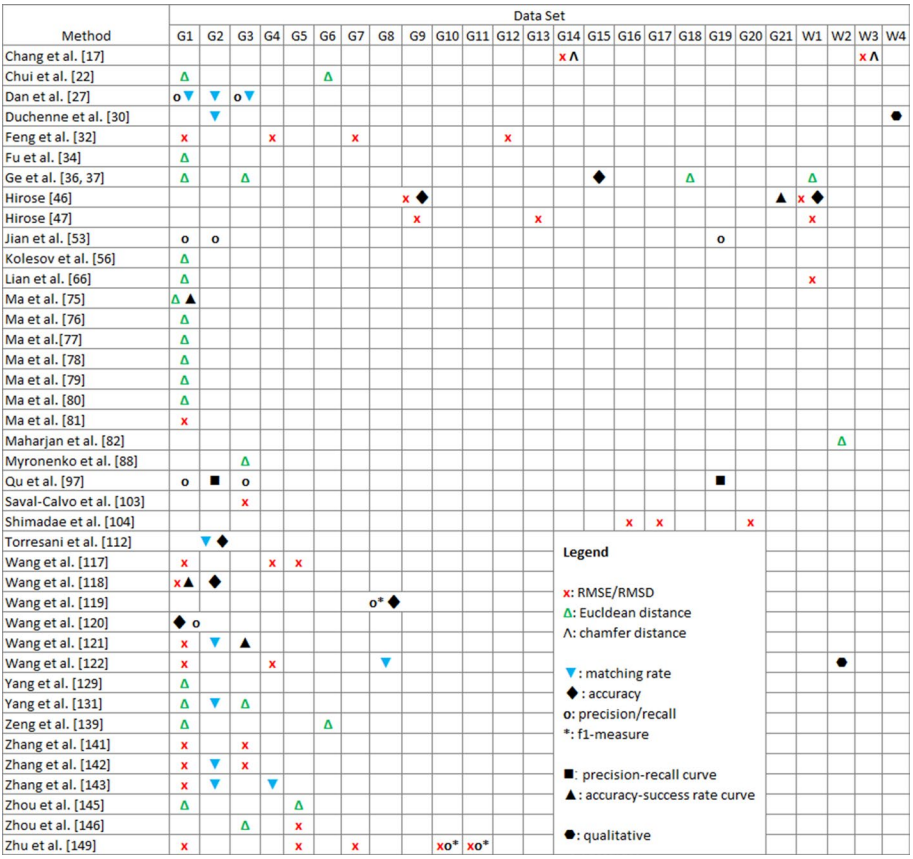


**Fig. 5** Data sets and evaluation metrics used by the existing registration methods. The data sets are represented by their indices (see Table 4 for the data sets)

## 4 Conclusion, trend and future work

Non-rigid point set registration has been used in a wide range of computer vision applications and is a very challenging task. Methods have been developed for the challenges such as data degradation and deformation. This paper presents a review and analysis of the existing methods to gain a better understanding of the open challenges, existing strategies, and future trends.

A registration method optimizes an objective function including constraints, which ensures well-posed, incorporates prior knowledge, and deals with data degradation. These constraints are employed to limit the search space for correspondence and transformation functions to achieve a plausible registration result. Besides model-based methods, graph matching techniques focus on finding the correspondences by embedding points into a graph and leveraging graph properties in search of a solution. The graph matching methods, however, are intractable and typically handled using approximation solutions (Chang et al. 2020; Leordeanu and Hebert 2005). As deep neural networks gain popularity in many fields and demonstrate superior performance in computer vision applications, learning-based methods have emerged and achieved promising

results in registration problems. Yet, the challenges of the deep learning methods lie in the demand for a large amount of training data to learn the parameters of the network and limited generalization power to various applications.

Integrating prior knowledge includes incorporating prior correspondence (or correspondence weight) value, using additional information (such as color information of points) while computing correspondences, adding constraints in the deformation model, and implementing regularization techniques. Data degradation is the major cause of problems in point set registration. Ideas to overcome data degradation include enforcing smoothness to the transformation function, preserving local neighborhood structure, rejecting incorrect correspondences from initial putative correspondences, and explicitly adding additional uniform distribution. These techniques have been seamlessly integrated into model-based methods as well as learning-based methods.

Although the existing methods have achieved significant progress, non-rigid point set registration remains a major research focus with many open technical challenges. The following lists a few of the most prominent ones:

- Identifying landmark correspondence plays an important role in the large and uneven deformation. Existing methods typically extract landmarks and establish the correspondence of landmarks to assist in matching the rest of the points. Finding landmarks is easier than finding the correct landmark correspondences between the point sets because the search space of landmark correspondence matching is huge. Currently, techniques such as using higher-order graph matching (Chang et al. 2020; Zeng et al. 2016), manually established landmark correspondences (Maharjan et al. 2021), or relying on feature descriptors matching (Feng and Feng 2020) are used to deal with this challenge. Future research should explore computing efficient and reliable landmark correspondences that work for various objects.
- Many low-cost devices including mobile devices capture depth images. Registration becomes much more difficult when the data are captured from different types of devices. For example, when the resolution of the point sets is different, the registration methods need to deal with challenging scenarios such as missing correspondences and point sets with different sizes. Some studies explored cross-source point cloud registration in rigid case (Huang et al. 2017), while others studied group-wise registration of the point sets, which are obtained from similar devices (Wang et al. 2008). Further research is needed to explore registering multiple point sets (i.e., group-wise registration).
- Evaluating the robustness of data degradation is very important to understand the effectiveness of a registration method. There are data sets, e.g., 3D Face (Myronenko and Song 2010), that provide different levels of deformation, but include no differentiation among the magnitudes of noise, outliers, and occlusions. Such levels of degradation are absent from the majority of the public data sets. Many studies prepare different levels of data degradation in the experiments but lack consistency in rating the deformation among the data sets. It is critical to systematically generate quantitative grades for the magnitude of deformation.
- Speed of registration is a bottleneck for applying the existing methods to real-world applications, especially when the size of the point sets is large. Downsampling is often used, which reduces the processing time but is still insufficient for achieving a balanced result of real-time performance and satisfactory accuracy. Deep network-based methods demonstrate promising performance on the registration speed. However, the major challenges faced by such methods include the inferior generalization ability to multiple categories, high computational costs, and registration accuracy.

**Table 5** Registration performance in terms of RMSE and Euclidean distance

| Method | Data | Metric | Level | Error |
|---|---|---|---|---|
| Ma et al. (2016) | 3D Wolf Shape (Kim et al. 2011) | RMSE | 1 | 0.50, 0.74, 0.58 |
| (PR-GLS) | | | 2 | 0.71, 0.92, 0.45 |
| Ma et al. (2019a) | Fish (Chui and Rangarajan 2000b, 2003) | RMSE | 1 | 0.0027 |
| (MR-RPM) | | | 2 | 0.0047 |
| | | | 3 | 0.0068 |
| | | | 4 | 0.0106 |
| | | | 5 | 0.0131 |
| | Chinese Char. (Chui and Rangarajan 2000b, 2003) | RMSE | 1 | 0.0056 |
| | | | 2 | 0.0092 |
| | | | 3 | 0.0142 |
| | | | 4 | 0.0195 |
| | | | 5 | 0.0222 |
| Zeng et al. (2017b) | Retinal Images | RMSE | | 3.08 |
| (GL-CATE) | BIT (BIT: biometrics ideal test, http://biometrics.idealtest.org/) | RMSE | | 5.34 |
| | Remote Sensing Images | RMSE | | 0.88 |
| Zhou et al. (2018) | 4DCT | Euclidean distance | | 0.455 (mm) |
| (DSMM) | COPD | Euclidean distance | | 0.917 (mm) |
| Yang et al. (2018) | MNIST Handwritten (LeCun et al. 1998) | Euclidean distance | | 0.0229 |
| | MPEG-7 shape silhouette (Jeannin and Bober 1999) | Euclidean distance | | 0.0316 |

**Table 6** Registration performance in terms of precision and recall

| Method | Data | Precision | Recall |
|---|---|---|---|
| Wang et al. (2016a) | Fish (Chui and Rangarajan 2003) | 99.33 | 98.88 |
| (MRCVF) | Chinese Character (Chui and Rangarajan 2003) | 98.84 | 96.53 |
| | The Oxford Buildings (Philbin et al. 2007) | 95.94 | 99.70* |
| Ma et al. (2014) | Affine Covariant Regions (Mikolajczyk et al. 2005) | 98.57 | 97.75 |
| (VFC) | DogCat | 100.00 | 100.00 |
| | T-shirt | 90.67 | 93.15 |
| | Mex (Tuytelaars and Van Gool 2004) | 96.47 | 100.0 |
| | Tree (Tuytelaars and Van Gool 2004) | 94.85 | 97.87 |
| Ma et al. (2017) (LPM) | Affine Covariant Regions (Mikolajczyk et al. 2005) | 99.16 | 99.36* |

The ones with an asterisk * are the average time of multiple cases

**Data availibility** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

## References


3DScan: 3D Scanning Library, http://graphics.stanford.edu/data/3Dscanrep

Anguelov D, Srinivasan P, Koller D, Thrun S, Rodgers J, Davis J (2005) Scape: shape completion and animation of people. ACM Trans Graph 24(3):408–416

Archive3D: Archive3d, http://archive3d.net/

Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. In: Leonardis A, Bischof H, Pinz A (eds) Computer vision—ECCV 2006. Berlin, Heidelberg, pp 404–417

Bednarik J, Fua P, Salzmann M (2018) Learning to reconstruct texture-less deformable surfaces from a single view. In: 2018 international conference on 3d vision (3DV), pp 606–615

Belkin M, Niyogi P (2008) Towards a theoretical foundation for Laplacian-based manifold methods. J Comput Syst Sci 74(8):1289–1308 (**learning theory 2005**)

Belkin M, Niyogi P, Sindhwani V (2006) Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. J Mach Learn Res 7:2399–2434

Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 24(4):509–522

BIT: biometrics ideal test, http://biometrics.idealtest.org/

Bogo F, Romero J, Pons-Moll G, Black MJ (2017) Dynamic faust: registering human bodies in motion. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 5573–5582

Bronstein A, Bronstein M, Kimmel R (2008a) Numerical geometry of non-rigid shapes. Springer, Incorporated

Bronstein AM, Bronstein MM, Bruckstein AM, Kimmel R (2008b) Analysis of two-dimensional non-rigid shapes. Int J Comput Vis 78(1):67–88

Cao H, Wang H, Zhang N, Yang Y, Zhou Z (2022) Robust probability model based on variational bayes for point set registration. Knowl Based Syst 241:108182

Castillo R Dir-lab, https://www.dir-lab.com/index.html

CCC: complicated chinese characters, https://github.com/xdregis/complicated_chinese_characters

Chang S, Ahn C, Lee M, Oh S (2020) Graph-matching-based correspondence search for nonrigid point cloud registration. Comput Vis Image Underst 192:102899

Chen Z, Haykin S (2002) On different facets of regularization theory. Neural Comput 14(12):2791–2846

Chen QY, Feng DZ, Hu HS (2022) A robust non-rigid point set registration algorithm using both local and global constraints. Vis Comput

Cho M, Alahari K, Ponce J (2013) Learning graphs to match. In: 2013 IEEE international conference on computer vision. pp 25–32

Chui H, Rangarajan A (2000a) A feature registration framework using mixture models. In: Proceedings IEEE workshop on mathematical methods in biomedical image analysis. MMBIA-2000 (Cat. No.PR00737). pp 190–197
Chui H, Rangarajan A (2000b) A new algorithm for non-rigid point matching. In: Proceedings IEEE conference on computer vision and pattern recognition. CVPR 2000 (Cat. No.PR00662) 2:44–51
Chui H, Rangarajan A (2003) A new point matching algorithm for non-rigid registration. Comput Vis Image Underst 89(2):114–141
Collet A, Chuang M, Sweeney P, Gillett D, Evseev D, Calabrese D, Hoppe H, Kirk A, Sullivan S (2015) High-quality streamable free-viewpoint video. ACM Trans Graph 34(4):1–3
Combes B, Prima S (2020) An efficient EM-ICP algorithm for non-linear registration of large 3D point sets. Comput Vis Image Underst 191:102854
Cour T, Srinivasan P, Shi J (2007) Balanced graph matching. In: Schölkopf B, Platt J, Hoffman T (eds) Advances in neural information processing systems, vol 19. MIT Press, Cambridge
Dan T, Yang Y, Xing L, Yang K, Zhang Y, Ong SH, Song F, Gao X (2018) Multifeature energy optimization framework and parameter adjustment-based nonrigid point set registration. J Appl Remote Sens 12(3):1–27
Datasets C Cmu house and hotel sequence images, http://vasc.ri.cmu.edu/idb/html/motion/house/index.html
Duchenne O, Bach F, Kweon IS, Ponce J (2009) A tensor-based algorithm for high-order graph matching. In: 2009 IEEE conference on computer vision and pattern recognition. pp 1980–1987
Duchenne O, Bach F, Kweon IS, Ponce J (2011) A tensor-based algorithm for high-order graph matching. IEEE Trans Pattern Anal Mach Intell 33(12):2383–2395
Dupej J, Vaclav Pelikan J (2015) Low-rank matrix approximations for coherent point drift. Pattern Recogn Lett 52:53–58
Feng XW, Feng DZ (2020) A robust nonrigid point set registration method based on collaborative correspondences. Sensors (Basel) 20:3248
Feng W, Zhang J, Cai H, Xu H, Hou J, Bao H (2021) Recurrent multi-view alignment network for unsupervised surface registration. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). pp 10292–10302
Frome A, Huber D, Kolluri R, Bülow T, Malik J (2004) Recognizing objects in range data using regional point descriptors. Computer vision - ECCV 2004. Springer, Berlin, Heidelberg, pp 224–237
Fu M, Zhou W (2016) Non-rigid point set registration via mixture of asymmetric gaussians with integrated local structures. In: 2016 IEEE international conference on robotics and biomimetics (ROBIO). pp 999–1004
Ge S, Fan G (2015) Non-rigid articulated point set registration with local structure preservation. In: 2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW). pp 126–133
Ge S, Fan G (2019) Topology-aware non-rigid point set registration via global—local topology preservation. Mach Vis Appl 30(4):717–735
Ge S, Fan G, Ding M (2014) Non-rigid point set registration with global-local topology preservation. In: 2014 IEEE conference on computer vision and pattern recognition workshops. pp 245–251
Glaunes J, Trouvé A, Younes L (2004) Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and sub-manifolds matching. In: Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition. pp 712–718. CVPR'04, IEEE Computer Society, USA
Golyanik V, Taetz B, Reis G, Stricker D (2016) Extended coherent point drift algorithm with correspondence priors and optimal subsampling. In: 2016 IEEE winter conference on applications of computer vision (WACV). pp 1–9
Golyanik V, Shimada S, Varanasi K, Stricker D (2018) Hdm-net: monocular non-rigid 3D reconstruction with learned deformation model. In: Bourdot P, Cobb S, Interrante V, kato H, Stricker D (eds) Virtual reality and augmented reality. Springer, Cham, pp 51–72
Greengard L, Strain J (1991) The fast Gauss transform. SIAM J Sci Comput 12(1):79–94
Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. Springer Series in Statistics. Springer, New York
Hein M, Audibert JY, von Luxburg U (2005) From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians. Learning theory. Springer, Berlin, Heidelberg, pp 470–485
Hinton GE, Williams CKI, Revow MD (1992) Adaptive elastic models for hand-printed character recognition. In: Moody JE, Hanson SJ, Lippmann RP (eds) Advances in neural information processing systems. Morgan-Kaufmann, Burlington, pp 512–519
Hirose O (2017) Dependent landmark drift: robust point set registration with a gaussian mixture model and a statistical shape model. ArXiv abs/1711.06588

Hirose O (2020) Acceleration of non-rigid point set registration with downsampling and gaussian process regression. IEEE Trans Pattern Anal Mach Intell 43(8):2858–2865
Hirose O (2021) A Bayesian formulation of coherent point drift. IEEE Trans Pattern Anal Mach Intell 43(7):2269–2286
Horaud R, Niskanen M, Dewaele G, Boyer E (2009) Human motion tracking by registering an articulated surface to 3D points and normals. IEEE Trans Pattern Anal Mach Intell 31(1):158–163
Huang X, Zhang J, Fan L, Wu Q, Yuan C (2017) A systematic approach for cross-source point cloud registration by preserving macro and micro structures. IEEE Trans Image Process 26(7):3261–3276
Huang X, Mei G, Zhang J (2020) Feature-metric registration: a fast semi-supervised approach for robust point cloud registration without correspondences. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). pp 11363–11371
Jeannin S, Bober M (1999) Description of core experiments for mpeg-7 motion/shape. MPEG-7, ISO/IEC/JTC1/SC29/WG11/MPEG99 N 2690
Jian B, Vemuri BC (2005) A robust algorithm for point set registration using mixture of gaussians. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. 2:1246–1251
Jian B, Vemuri BC (2011) Robust point set registration using gaussian mixture models. IEEE Trans Pattern Anal Mach Intell 33(8):1633–1645
Johnson AE, Hebert M (1999) Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Trans Pattern Anal Mach Intell 21(5):433–449
Kim VG, Lipman Y, Funkhouser T (2011) Blended intrinsic maps. In: ACM SIGGRAPH 2011 Papers. SIGGRAPH '11, Association for Computing Machinery, New York, NY, USA
Kolesov I, Lee J, Vela P, Tannenbaum A (2013a) A stochastic approach for non-rigid image registration. In: Egiazarian KO, Agaian SS, Gotchev AP (eds) Image processing: algorithms and systems XI, vol 8655. International Society for Optics and Photonics, SPIE, Bellingham, pp 282–296
Kolesov I, Lee J, Vela P, Tannenbaum A (2013b) Stochastic image registration with user constraints. In: Ourselin S, Haynor DR (eds) Medical imaging 2013: image processing, vol 8669. International Society for Optics and Photonics, SPIE, Bellingham, pp 845–851
Kolesov I, Lee J, Sharp G, Vela P, Tannenbaum A (2016) A stochastic approach to diffeomorphic point set registration with landmark constraints. IEEE Trans Pattern Anal Mach Intell 38(2):238–251
Kong L, Yuan X, Maharjan AM (2018) A hybrid framework for automatic joint detection of human poses in depth frames. Pattern Recogn 77:216–225
LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
Leordeanu M, Hebert M (2005) A spectral technique for correspondence problems using pairwise constraints. In: Tenth IEEE international conference on computer vision (ICCV'05) Volume 1. 2:1482–1489
Li Y, Harada T (2022) Lepard: learning partial point cloud matching in rigid and deformable scenes. In: 2022 IEEE/cvf conference on computer vision and pattern recognition (CVPR). pp 5554–5564
Li T, Bolkart T, Black MJ, Li H, Romero J (2017) Learning a model of facial shape and expression from 4D scans. ACM Trans Graph 36(6):194
Li X, Wang L, Fang Y (2019) Pc-net: Unsupervised point correspondence learning with neural networks. In: 2019 international conference on 3D vision (3DV). pp 145–154
Lian W, Zhang L, Yang MH (2017) An efficient globally optimal algorithm for asymmetric point matching. IEEE Trans Pattern Anal Mach Intell 39(7):1281–1293
Ling H, Jacobs DW (2007) Shape classification using the inner-distance. IEEE Trans Pattern Anal Mach Intell 29(2):286–299
Lipman Y, Funkhouser T (2009) Möbius voting for surface correspondence. ACM Trans Graph 28(3):1–2
Loiola EM, de Abreu NMM, Boaventura-Netto PO, Hahn P, Querido T (2007) A survey for the quadratic assignment problem. Eur J Oper Res 176(2):657–690
Lowe DG (1999) Object recognition from local scale-invariant features. In: Proceedings of the seventh IEEE international conference on computer vision. 2: 1150–1157
Lu Q, Lu Y, Xiao M, Yuan X, Jia W (2019a) 3D-Fhnet: three-dimensional fusion hierarchical reconstruction method for any number of views. IEEE Access 7:172902–172912
Lu Q, Xiao M, Lu Y, Yuan X, Yu Y (2019b) Attention-based dense point cloud reconstruction from a single image. IEEE Access 7:137420–137431
Lu W, Wan G, Zhou Y, Fu X, Yuan P, Song S (2019c) Deepvcp: an end-to-end deep neural network for point cloud registration. In: 2019 IEEE/CVF international conference on computer vision (ICCV). pp 12–21
Luo J, Yuan M, Fu K, Wang M, Zhang C (2022) Deep graph matching based dense correspondence learning between non-rigid point clouds. IEEE Robot Autom Lett 7(3):5842–5849

Ma J, Zhao J, Tian J, Tu Z, Yuille AL (2013) Robust estimation of nonrigid transformation for point set registration. In: 2013 IEEE conference on computer vision and pattern recognition. pp 2147–2154
Ma J, Zhao J, Tian J, Yuille AL, Tu Z (2014) Robust point matching via vector field consensus. IEEE Trans Image Process 23(4):1706–1721
Ma J, Qiu W, Zhao J, Ma Y, Yuille AL, Tu Z (2015) Robust $l_2e$ estimation of transformation for non-rigid registration. IEEE Trans Signal Process 63(5):1115–1129
Ma J, Zhao J, Yuille AL (2016) Non-rigid point set registration by preserving global and local structures. IEEE Trans Image Process 25(1):53–64
Ma J, Zhao J, Guo H, Jiang J, Zhou H, Gao Y (2017) Locality preserving matching. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence, IJCAI-17. pp 4492–4498
Ma J, Wu J, Zhao J, Jiang J, Zhou H, Sheng QZ (2019a) Nonrigid point set registration with robust transformation learning under manifold regularization. IEEE Trans Neural Netw Learn Syst 30(12):3584–3597
Ma J, Zhao J, Jiang J, Zhou H, Guo X (2019b) Locality preserving matching. Int J Comput Vision 127(5):512–531
Maharjan A, Yuan X (2020) Point set registration of large deformation using auxiliary landmarks. Urban intelligence and applications. Springer, Singapore, pp 86–98
Maharjan A, Yuan X (2022) Registration of human point set using automatic key point detection and region-aware features. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp 741–749
Maharjan A, Yuan X, Lu Q, Fan Y, Chen T (2021) Non-rigid registration of point clouds using landmarks and stochastic neighbor embedding. J Electron Imaging 30(3):1–15
Maiseli B, Gu Y, Gao H (2017) Recent developments and trends in point set registration methods. J Vis Commun Image Represent 46:95–106
Marin R, Melzi S, Rodolà E, Castellani U (2020) Farm: functional automatic registration method for 3D human bodies. Comput Graph Forum 39(1):160–173
Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T, Gool LV (2005) A comparison of affine region detectors. Int J Comput Vision 65:2005
Myronenko A, Song X (2010) Point set registration: coherent point drift. IEEE Trans Pattern Anal Mach Intell 32(12):2262–2275
Myronenko A, Song X, Nán MACP (2007) Non-rigid point set registration: coherent point drift. In: Schölkopf B, Platt JC, Hoffman T (eds) Advances in neural information processing systems 19. MIT Press, Cambridge, pp 1009–1016
Nordstrøm MM, Larsen M, Sierakowski J, Stegmann MB (2004) The IMM face database—an annotated dataset of 240 face images. Tech. rep., Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby
OSU A Accad motion capture data, https://accad.osu.edu/research/motion-lab/mocap-system-and-data
Pais GD, Miraldo P, Ramalingam S, Govindu VM, Nascimento JC, Chellappa R (2020) 3DRegNet: a deep neural network for 3D point registration. pp 7191–7201
Panaganti V, Aravind R (2015) Robust nonrigid point set registration using graph-laplacian regularization. In: IEEE winter conference on applications of computer vision. pp 1137–1144
Papadimitriou CH, Steiglitz K (1982) Combinatorial optimization: algorithms and complexity. Prentice-Hall Inc, USA
Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2007) Object retrieval with large vocabularies and fast spatial matching. In: IEEE conference on computer vision and pattern recognition
Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2008) Lost in quantization: improving particular object retrieval in large scale image databases. In: IEEE conference on computer vision and pattern recognition
Qu HB, Wang JQ, Li B, Yu M (2017) Probabilistic model for robust affine and non-rigid point set matching. IEEE Trans Pattern Anal Mach Intell 39(2):371–384
Rangarajan A, Chui H, Mjolsness E, Pappu S, Davachi L, Goldman-Rakic P, Duncan J (1997) A robust point-matching algorithm for autoradiograph alignment. Med Image Anal 1:379–398
Revow M, Williams CKI, Hinton GE (1996) Using generative models for handwritten digit recognition. IEEE Trans Pattern Anal Mach Intell 18(6):592–606
Rusu RB, Cousins S (2011) 3D is here: point cloud library (PCL). In: IEEE international conference on robotics and automation (ICRA). Shanghai, China
Rusu RB, Blodow N, Marton ZC, Beetz M (2008) Aligning point cloud views using persistent feature histograms. In: 2008 IEEE/RSJ international conference on intelligent robots and systems. pp 3384–3391
Rusu RB, Blodow N, Beetz M (2009) Fast point feature histograms (fpfh) for 3d registration. In: 2009 IEEE international conference on robotics and automation. pp 3212–3217

Saval-Calvo M, Azorin-Lopez J, Fuster-Guillo A, Villena-Martinez V, Fisher RB (2018) 3D non-rigid registration using color: color coherent point drift. Comput Vis Image Underst 169:119–135
Shimada S, Golyanik V, Tretschk E, Stricker D, Theobalt C (2019) Dispvoxnets: non-rigid point set alignment with supervised learning proxies. In: 2019 international conference on 3D vision (3DV). pp 27–36
Song G, Fan G (2015) Articulated non-rigid point set registration for human pose estimation from 3D sensors. Sensors (Basel, Switzerland) 15:15218–45
Stegmann MB, Gomez DD (2002) A brief introduction to statistical shape analysis (mar 2002), images, annotations and data reports are placed in the enclosed zip-file
Sumner RW, Popović J (2004) Deformation transfer for triangle meshes. ACM Trans Graph 23(3):399–405
Tajdari F, Huysmans T, Yang Y, Song Y (2022) Feature preserving non-rigid iterative weighted closest point and semi-curvature registration. IEEE Trans Image Process 31:1841–1856
Tam GK, Cheng ZQ, Lai YK, Langbein FC, Liu Y, Marshall D, Martin RR, Sun XF, Rosin PL (2013) Registration of 3D point clouds and meshes: a survey from rigid to nonrigid. IEEE Trans Vis Comput Gr 19(7):1199–1217
Tao W, Sun K (2014) Asymmetrical gauss mixture models for point sets matching. In: 2014 IEEE conference on computer vision and pattern recognition. pp 1598–1605
Torresani L, Kolmogorov V, Rother C (2008) Feature correspondence via graph matching: models and global optimization. In: Forsyth D, Torr P, Zisserman A (eds) Computer vision—ECCV 2008. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 596–609
Torresani L, Kolmogorov V, Rother C (2013) A dual decomposition approach to feature correspondence. IEEE Trans Pattern Anal Mach Intell 35(2):259–271
Trappolini G, Cosmo L, Moschella L, Marin R, Melzi S, Rodola E (2021) Shape registration in the time of transformers. In: 35th conference on neural information processing systems (NeurIPS 2021)
Tsin Y, Kanade T (2004) A correlation-based approach to robust point set registration. In: Pajdla T, Matas J (eds) Computer vision— ECCV 2004. Springer, Berlin, Heidelberg, pp 558–569
Tuytelaars T, Van Gool L (2004) Matching widely separated views based on affine invariant regions. Int J Comput Vis 59(1):61–85
Wahba G (1990) Spline models for observational data. Soc Ind Appl Math
Wang G, Chen Y (2017) Fuzzy correspondences guided gaussian mixture model for point set registration. Knowl Based Syst 136:200–209
Wang G, Chen Y (2021) Scm: spatially coherent matching with Gaussian field learning for nonrigid point set registration. IEEE Trans Neural Net Learn Syst 32(1):203–213
Wang L, Fang Y (2019) Coherent point drift networks: unsupervised learning of non-rigid point set registration. CoRR abs/1906.03039
Wang F, Vemuri BC, Rangarajan A, Eisenschenk SJ (2008) Simultaneous nonrigid registration of multiple point sets and atlas construction. IEEE Trans Pattern Anal Mach Intell 30(11):2011–2022
Wang G, Wang Z, Chen Y, Liu X, Ren Y, Peng L (2016a) Learning coherent vector fields for robust point matching under manifold regularization. Neurocomputing 216:393–401
Wang H, Liu X, Yuan X, Liang D (2016b) Multi-perspective terrestrial LiDAR point cloud registration using planar primitives. In: 2016 IEEE international geoscience and remote sensing symposium (IGARSS). pp 6722–6725
Wang G, Zhou Q, Chen Y (2017) Robust non-rigid point set registration using spatially constrained gaussian fields. IEEE Trans Image Process 26(4):1759–1769
Wang G, Chen Y, Zheng X (2018) Gaussian field consensus: a robust nonparametric matching method for outlier rejection. Pattern Recogn 74:305–316
Wang Z, Xue N, Lei L, Xia GS (2022) Partial wasserstein adversarial network for non-rigid point set registration. In: The international conference on learning representations (ICLR)
Williams C, Seeger M (2001) Using the nyström method to speed up kernel machines. In: Leen T, Dietterich T, Tresp V (eds) Advances in neural information processing systems, vol 13. MIT Press, Cambridge
Yang J (2011) The thin plate spline robust point matching (tps-rpm) algorithm: A revisit. Pattern Recogn Lett 32(7):910–918
Yang C, Duraiswami R, Gumerov NA, Davis L (2003) Improved fast gauss transform and efficient kernel density estimation. In: Proceedings ninth IEEE international conference on computer vision. 1:664–671
Yang Y, Ong SH, Foong KWC (2015) A robust global and local mixture distance based non-rigid point set registration. Pattern Recogn 48(1):156–173
Yang C, Liu Y, Jiang X, Zhang Z, Wei L, Lai T, Chen R (2018) Non-rigid point set registration via adaptive weighted objective function. IEEE Access 6:75947–75960

Ye M, Shen Y, Du C, Pan Z, Yang R (2016) Real-time simultaneous pose and shape estimation for articulated objects using a single depth camera. IEEE Trans Pattern Anal Mach Intell 38(8):1517–1532
Yefeng Zheng, Doermann D (2006) Robust point matching for nonrigid shapes by preserving local neighborhood structures. IEEE Trans Pattern Anal Mach Intell 28(4):643–649
Yuan X, Kong L, Feng D, Wei Z (2017) Automatic feature point detection and tracking of human actions in time-of-flight videos. IEEE/CAA J Autom Sin 4(4):677–685
Yuan X, Feng D, Zuo Z (2018) Automatic construction of aerial corridor from discrete LiDAR point cloud. Springer, Cham, pp 449–465
Yuille AL, Grzywacz NM (1988) The motion coherence theory. In: [1988 Proceedings] second international conference on computer vision. pp 344–353
Zaharescu A, Boyer E, Varanasi K, Horaud R (2009) Surface feature detection and description with applications to mesh matching. In: 2009 IEEE conference on computer vision and pattern recognition. pp 373–380
Zeng Y, Wang C, Wang Y, Gu X, Samaras D, Paragios N (2010) Dense non-rigid surface registration using high-order graph matching. In: 2010 IEEE computer society conference on computer vision and pattern recognition. pp 382–389
Zeng Y, Wang C, Gu X, Samaras D, Paragios N (2016) Higher-order graph principles towards non-rigid surface registration. IEEE Trans Pattern Anal Mach Intell 38(12):2416–2429
Zeng Y, Qian Y, Zhu Z, Hou J, Yuan H, He Y (2021) Corrnet3D: unsupervised end-to-end learning of dense correspondence for 3D point clouds. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). pp 6048–6057
Zhang L, Snavely N, Curless B, Seitz SM (2004) Spacetime faces: high resolution capture for modeling and animation. ACM Trans Graph 23(3):548–558
Zhang P, Qiao Y, Wang S, Yang J, Zhu Y (2017a) A robust coherent point drift approach based on rotation invariant shape context. Neurocomputing 219:455–473
Zhang S, Yang Y, Yang K, Luo Y, Ong SH (2017b) Point set registration with global-local correspondence and transformation estimation. In: 2017 IEEE international conference on computer vision (ICCV). pp 2688–2696
Zhong Y (2009) Intrinsic shape signatures: A shape descriptor for 3d object recognition. In: 2009 IEEE 12th international conference on computer vision workshops, ICCV workshops. pp 689–696
Zhou Z, Zheng J, Dai Y, Zhou Z, Chen S (2014) Robust non-rigid point set registration using student's-t mixture model. PLoS ONE 9:e91381
Zhou Z, Tong B, Geng C, Hu J, Zheng J, Dai Y (2017) Direct point-based registration for precise non-rigid surface matching using student's-t mixture model. Biomed Signal Process Control 33:10–18
Zhou Z, Tu J, Geng C, Hu J, Tong B, Ji J, Dai Y (2018) Accurate and robust non-rigid point set registration using student's-t mixture model with prior probability modeling. Sci Rep 8:1–7
Zhu H, Guo B, Zou K, Li Y, Yuen KV, Mihaylova L, Leung H (2019a) A review of point set registration: from pairwise registration to groupwise registration. Sensors 19(5):1191
Zhu H, Zou K, Li Y, Cen M, Mihaylova L (2019b) Robust non-rigid feature matching for image registration using geometry preserving. Sensors 19(12):2729